

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:16:19 ; Search time 140.424 Seconds  
(without alignments)  
2428.962 Million cell updates/sec

Title: US-10-773-302-2  
Perfect score: 4051  
Sequence: 1 MGQYKSKSEKSPVCRKRVK.....RYSQADALKYIGIEREMRIP 746

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*
- 10: geneseqp2006s:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4051	100.0	746	2	AAW05260 Chromatin
2	4038	99.7	746	6	ABU56461 Lung canc
3	4038	99.7	746	7	ADG10642 Human STA
4	4038	99.7	746	7	ADN39392 Cancer/an
5	4030.5	99.5	751	8	ADP95920 Human EZH
6	3812.5	94.1	707	7	ADP65008 Human pro
7	2845	70.2	517	7	ADG10764 Human STA
8	2601.5	64.2	747	9	ADX05848 Cycilin-de
9	2429.5	60.0	723	6	ABU11057 Human pro
10	2156.5	53.2	760	4	ABB62427 Drosophi
11	829	20.5	895	8	ADT57579 Plant pol
12	828.5	20.5	880	8	ADY25049 Plant ful
13	810	20.0	893	5	AAO14530 Protein o
14	805	19.9	891	8	ADM48181 Polypepti
15	795	19.6	931	8	ADT57263 Plant pol
16	783	19.3	921	3	AAV32297 Corn poly
17	781	19.3	931	5	AAO14529 Protein o
18	731	18.0	509	8	ADX94792 Plant ful
19	726	17.9	689	2	AAV42698 Arabidops
20	726	17.9	689	3	AAV57036 Petillisa
21	726	17.9	689	4	AAH82455 Arabidops
22	726	17.9	689	6	ABP70523 Histone d
23	726	17.9	932	6	ABP70537 Amino aci

24	725	17.9	689	3	AAH01673 FISI prot
25	719.5	17.8	463	8	ADX94788 Plant ful
26	713	17.6	417	8	ADY95305 Plant ful
27	700.5	17.3	536	8	ADY13472 Plant ful
28	681.5	16.8	346	8	ADX94796 Plant ful
29	659	16.3	294	8	ADY14004 Plant ful
30	643.5	15.9	298	8	ADX93483 Plant ful
31	643.5	15.9	326	8	ADY65894 Plant ful
32	631	15.6	196	9	ADY65574 S. manson
33	625	15.4	256	8	ADX71472 Plant ful
34	622.5	15.4	244	8	ADX94793 Plant ful
35	582.5	14.4	225	8	ADX94797 Plant ful
36	576.5	14.2	359	3	AAV32298 Wheat pol
37	474	11.7	114	10	ABE98254 Human EZH
38	473.5	11.7	93	8	ADG36757 PR/SET do
39	430	10.6	153	8	ADX71195 Plant ful
40	409	10.1	102	4	AAH15631 Peptide #
41	409	10.1	102	4	ABB34636 Peptide #
42	409	10.1	102	4	AAH28129 Peptide #
43	409	10.1	102	4	ABH29459 Peptide #
44	409	10.1	102	4	AAH67821 Human Don
45	409	10.1	102	4	AAH03374 Peptide #

## ALIGNMENTS

RESULT 1  
AAW05260 standard; protein; 746 AA.

AAW05260; 25-MAR-2003 (revised)  
DT 05-MAY-1997 (first entry)  
DE Chromatin regulator protein EZH2.  
KW Chromatin; regulator; EZH1; EZH2; SUV39H; SET domain; gene therapy;  
KW Cancer.  
XX Homo sapiens.  
XX OS  
XX W09635784-A2.  
XX 14-NOV-1996.  
XX 02-MAY-1996; 96WO-EP001818.  
XX 10-MAY-1995; 95DR-01016776.  
XX (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
XX Jenuwein T, Laible G;  
XX WPI; 1996-518672/51.  
XX N-PSDB; AAT43624.  
XX New DNA encoding chromatin regulator protein with SET domain - and  
XX related vectors, transformed cells, proteins and antibodies, for  
XX diagnosis and treatment of cancer.  
XX Claim 12; Fig 7; 38pp; German.  
XX The DNA was isolated by screening a human B cell cDNA library with mixed  
XX Drosophila DNA probes based on the conserved SET domains in B(z) and  
XX Su(var)-3-9. The DNA, and its products, are useful in therapy (esp. gene  
XX therapy) and diagnosis of human diseases that involve deregulated  
XX chromatin-regulator genes having a SET domain, esp. cancer. (Updated on  
XX 25-MAR-2003 to correct PR field.)  
XX Sequence 746 AA;  
XX SQ

Query Match 100.0%; Score 4051; DB 2; Length 746;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-303;  
 Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGOTGKSEKGPVCMRRKRVKSEYMRRLQKRFPRADVKSMFSSNRKILERTILLNQEW 60
DB 1 MGOTGKSEKGPVCMRRKRVKSEYMRRLQKRFPRADVKSMFSSNRKILERTILLNQEW 60
QY 61 KORRIQPHILTSVSLRGRECSVTSDDLPPQVTPILKTLNVAASVPIMYSPLQONF 120
DB 61 KORRIQPHILTSVSLRGRECSVTSDDLPPQVTPILKTLNVAASVPIMYSPLQONF 120
QY 121 MVEDETVLHNI.PYMGDEVLDODGTFIEELIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180
DB 121 MVEDETVLHNI.PYMGDEVLDODGTFIEELIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180
QY 181 YNDDDDDDDDPPEEREKQDLJEDHRDKESRPPRKPSPDKIFEAISSMFPDGTAEEL 240
DB 181 YNDDDDDDDDPPEEREKQDLJEDHRDKESRPPRKPSPDKIFEAISSMFPDGTAEEL 240
QY 241 KEKYKELTEQOLPGALPPECTPNTIDGPNKSVOREOISLHSFHTLFCRCFCYDCFLHPFH 300
DB 241 KEKYKELTEQOLPGALPPECTPNTIDGPNKSVOREOISLHSFHTLFCRCFCYDCFLHPFH 300
QY 301 ATPNTYRKNTETALDNKPCGPOCYQHIEGAKERPAALATARITKTPPRPGRRGRULPN 360
DB 301 ATPNTYRKNTETALDNKPCGPOCYQHIEGAKERPAALATARITKTPPRPGRRGRULPN 360
QY 361 NSSRPFPTINVLKSDTSDREAGTETGGENNDKEBEKODETSSSEANSRCOTPIKM 420
DB 361 NSSRPFPTINVLKSDTSDREAGTETGGENNDKEBEKODETSSSEANSRCOTPIKM 420
QY 421 KPNIEPENVMWSGAEASMFVULIGTYDNECAIARLIGITCRQVYEFVKESSIIAPA 480
DB 421 KPNIEPENVMWSGAEASMFVULIGTYDNECAIARLIGITCRQVYEFVKESSIIAPA 480
QY 481 PAEVDTPPRKKRKHRLMAHCKRIQLKRGSSNHYNYNPOCDHPRQPCDSSGCPVIAQ 540
DB 481 PAEVDTPPRKKRKHRLMAHCKRIQLKRGSSNHYNYNPOCDHPRQPCDSSGCPVIAQ 540
QY 541 NFCEKFCOCCSECONRPFGCRCCKAQCNTKOCPCYIAVRECDPDLCTFGADHDSKNVS 600
DB 541 NFCEKFCOCCSECONRPFGCRCCKAQCNTKOCPCYIAVRECDPDLCTFGADHDSKNVS 600
QY 601 CKNCSIORGSKKHLILAPSIVAAGIFIKDPVQKNEFISEYCGEIIISQDEADRGKYVDK 660
DB 601 CKNCSIORGSKKHLILAPSIVAAGIFIKDPVQKNEFISEYCGEIIISQDEADRGKYVDK 660
QY 661 YMGCFLEFLNANDFYVDATRKGNKIRFANHSVNPNCYACVMVNGDHRIGIFAKAALQTGE 720
DB 661 YMGCFLEFLNANDFYVDATRKGNKIRFANHSVNPNCYACVMVNGDHRIGIFAKAALQTGE 720
QY 721 ELFPDYRSQADALKYVIGIEREMEIP 746
DB 721 ELFPDYRSQADALKYVIGIEREMEIP 746
  
```

OS Unidentified.  
 XX  
 PN WO200286443-A2.  
 XX 31-OCT-2002.  
 PD  
 XX 18-APR-2002; 2002MO-US012476.  
 PF  
 XX 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0339245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-0334370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 PA (BOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Aziz N, Murray R;  
 XX WP1; 2003-093161/08.  
 DR N-PSDB; ABX76182.  
 DR  
 XX  
 PT Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.  
 PS Claim 27; Page 232; 453pp; English.  
 XX  
 CC The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung cancer-  
 CC associated polynucleotides and polypeptides are used for identifying a  
 CC compound that modulates a lung cancer-associated polypeptide, for  
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
 CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell  
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 CC hyperinfectivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 CC for diagnostic purposes and as targets for screening for therapeutic  
 CC compounds that modulate lung cancer, such as antibodies. Sequences  
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the  
 CC invention  
 CC  
 XX  
 SQ Sequence 746 AA;  
 QY  
 DB 1 MGOTGKSEKGPVCMRRKRVKSEYMRRLQKRFPRADVKSMFSSNRKILERTILLNQEW 60  
 DB 1 MGOTGKSEKGPVCMRRKRVKSEYMRRLQKRFPRADVKSMFSSNRKILERTILLNQEW 60  
 QY 61 KORRIQPHILTSVSLRGRECSVTSDDLPPQVTPILKTLNVAASVPIMYSPLQONF 120  
 DB 61 KORRIQPHILTSVSLRGRECSVTSDDLPPQVTPILKTLNVAASVPIMYSPLQONF 120  
 QY 121 MVEDETVLHNI.PYMGDEVLDODGTFIEELIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180  
 DB 121 MVEDETVLHNI.PYMGDEVLDODGTFIEELIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180  
 QY 181 YNDDDDDDDDPPEEREKQDLJEDHRDKESRPPRKPSPDKIFEAISSMFPDGTAEEL 240  
 DB 181 YNDDDDDDDDPPEEREKQDLJEDHRDKESRPPRKPSPDKIFEAISSMFPDGTAEEL 240  
 QY 241 KEKYKELTEQOLPGALPPECTPNTIDGPNKSVOREOISLHSFHTLFCRCFCYDCFLHPFH 300  
 DB 241 KEKYKELTEQOLPGALPPECTPNTIDGPNKSVOREOISLHSFHTLFCRCFCYDCFLHPFH 300



Db 241 KEKYELTEQQLPGALPEECTPNIDGPNKSVQREOSLHSPHTLFCRCFKYDCFLHPH 300  
 QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 Db 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 QY 361 NSRPSTPTINVLKSDTDSREAGTENGNNDEKEBEKDETSSSSEANSRCOTPIKM 420  
 Db 361 NSRPSTPTINVLKSDTDSREAGTENGNNDEKEBEKDETSSSSEANSRCOTPIKM 420  
 QY 421 KENIEPPENVWMSGABSMFRVLIGTYYNFCALRLIGTKCRQYEFYRVESSTIAPA 480  
 Db 421 KENIEPPENVWMSGABSMFRVLIGTYYNFCALRLIGTKCRQYEFYRVESSTIAPA 480  
 QY 481 PAEDVDTPPRRKKRHRMAAHCRKIQLKDGSSNHNVTYQPCDHPROPCCSSCCCVIAQ 540  
 Db 481 PAEDVDTPPRRKKRHRMAAHCRKIQLKDGSSNHNVTYQPCDHPROPCCSSCCCVIAQ 540  
 QY 541 NFCEKPCQCSSECONRPFPCRCCKAQCNTKQPCYLAVERCDPDLCTGGAADHMSKNVS 600  
 Db 541 NFCEKPCQCSSECONRPFPCRCCKAQCNTKQPCYLAVERCDPDLCTGGAADHMSKNVS 600  
 QY 601 CNKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISQDEADRGKVVYDK 660  
 Db 601 CNKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISQDEADRGKVVYDK 660  
 QY 661 YWCSEFLPNLNDFFVDATRKGNKIRIPANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
 Db 661 YWCSEFLPNLNDFFVDATRKGNKIRIPANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
 QY 721 ELFPDYRYSOADALKVGIHERMEIP 746  
 Db 721 ELFPDYRYSOADALKVGIHERMEIP 746

## RESULT 3

ADG10642  
ID ADG10642 standard; protein; 746 AA.

AC ADG10642;

DT 26-FEB-2004 (first entry)

DE Human STAT6-activating protein, SEQ ID NO:232.

KW Human; STAT6-activating protein; drug screening; activator; inhibitor;  
 KW allergic disease; inflammation; autoimmune disease; diabetes;  
 KW hyperlipidemia; cancer; infection; HIV infection;  
 KW human immunodeficiency; cancer; Th1 hyperfunction; antiallergic;  
 KW antiinflammatory; antidiabetic; antilipemic; antiinfective; anti-HIV;  
 KW cytostatic; gene therapy; antisense therapy; ribozyme therapy.

OS Homo sapiens.

PN WO200296943-A1.

PD 05-DEC-2002.

PF 22-MAY-2002; 2002WO-JP004949.

PR 25-MAY-2001; 2001JP-00157043.

PR 30-AUG-2001; 2001JP-00260681.

PR 10-OCT-2001; 2001JP-00313175.

PA (ASAH ) ASAH KASEI KOGYO KK.

PI Honda G, Matsumoto A, Muramatsu S, Ishizawa K;

DR WPI; 2003-140442/13.

DR N-PSDB; ADG10641.

PT STAT6-activating proteins and encoded genes, applicable in diagnosis of  
and developing drugs to treat allergic diseases, inflammations,

PT autoimmune diseases, diabetes, hyperlipidemia, infections e.g. HIV, and  
 PT cancer.  
 PS Claim 1; SEQ ID NO 232; 2080bp; Japanese.  
 XX  
 CC The invention relates to 242 human STAT6-activating proteins and cDNAs  
 CC encoding them (ADG10411-ADG10894) and to sequences with 95% or more  
 CC homology to the STAT6-activating proteins and their encoding nucleic  
 CC acids. The invention also relates to recombinant vectors and host cells  
 CC comprising a STAT6-activating protein-encoding nucleic acid; the  
 CC recombinant production of a STAT6-activating protein; an antibody  
 CC specific for a STAT6-activating protein; antisense oligonucleotides and  
 CC ribozymes targeted to nucleic acids encoding a STAT6-activating protein;  
 CC methods of screening for activators or inhibitors of STAT6-activating  
 CC proteins; drug compositions comprising a modulator of STAT6-activating  
 CC protein activity or expression; and methods of treating patients by  
 CC administration of the drug compositions. The STAT6-activating proteins,  
 CC nucleic acids encoding them, and modulators of their activity or  
 CC expression are useful in the diagnosis and treatment of allergic  
 CC diseases, inflammation, autoimmune diseases, diabetes, hyperlipidemia,  
 CC cancer, infections (e.g., HIV), cancer and disorders associated with Th1  
 CC hyperfunction. The present sequence is related to the invention.

Sequence 746 AA;

Query Match 99.7%; Score 4038; DB 7; Length 746;

Best Local Similarity 99.7%; Pred. No. 9e-302; Mismatches 2; Indels 0; Gaps 0;

Db 1 MGQTKSEKGPVCKRKKVSEYMRRLQKFRRADEVKSMSSNRQKILERTETILNDEW 60  
 1 MGQTKSEKGPVCKRKKVSEYMRRLQKFRRADEVKSMSSNRQKILERTETILNDEW 60  
 Db 61 KORRIQPVHILTSVSLGTRGCVTSDDLDPPTQYIPLKTLNAVASVPIYMSWSPLOQNF 120  
 61 KORRIQPVHILTSVSLGTRGCVTSDDLDPPTQYIPLKTLNAVASVPIYMSWSPLOQNF 120  
 Db 121 MVEDETVLANIPYMGDEVLDDGTFIEELINYNQKVNQDRECGTINDEIPELVNALQ 180  
 121 MVEDETVLANIPYMGDEVLDDGTFIEELINYNQKVNQDRECGTINDEIPELVNALQ 180  
 Db 121 MVEDETVLANIPYMGDEVLDDGTFIEELINYNQKVNQDRECGTINDEIPELVNALQ 180  
 121 MVEDETVLANIPYMGDEVLDDGTFIEELINYNQKVNQDRECGTINDEIPELVNALQ 180  
 Db 181 YNDDDDDDGDDPEEREKOKLDEHRDKESRPPKRPSPDKIFPAISMPFDDKTAEBL 240  
 181 YNDDDDDDGDDPEEREKOKLDEHRDKESRPPKRPSPDKIFPAISMPFDDKTAEBL 240  
 Db 181 YNDDDDDDGDDPEEREKOKLDEHRDKESRPPKRPSPDKIFPAISMPFDDKTAEBL 240  
 181 YNDDDDDDGDDPEEREKOKLDEHRDKESRPPKRPSPDKIFPAISMPFDDKTAEBL 240  
 QY 241 KEKYELTEQQLPGALPEECTPNIDGPNKSVQREOSLHSPHTLFCRCFKYDCFLHPH 300  
 Db 241 KEKYELTEQQLPGALPEECTPNIDGPNKSVQREOSLHSPHTLFCRCFKYDCFLHPH 300  
 QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 Db 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 Db 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPAAALTARIKTPPRKPGRRGRLEPN 360  
 QY 361 NSRPSTPTINVLKSDTDSREAGTENGNNDEKEBEKDETSSSSEANSRCOTPIKM 420  
 Db 361 NSRPSTPTINVLKSDTDSREAGTENGNNDEKEBEKDETSSSSEANSRCOTPIKM 420  
 QY 421 KENIEPPENVWMSGABSMFRVLIGTYYNFCALRLIGTKCRQYEFYRVESSTIAPA 480  
 Db 421 KENIEPPENVWMSGABSMFRVLIGTYYNFCALRLIGTKCRQYEFYRVESSTIAPA 480  
 QY 481 PAEDVDTPPRRKKRHRMAAHCRKIQLKDGSSNHNVTYQPCDHPROPCCSSCCCVIAQ 540  
 Db 481 PAEDVDTPPRRKKRHRMAAHCRKIQLKDGSSNHNVTYQPCDHPROPCCSSCCCVIAQ 540  
 QY 541 NFCEKPCQCSSECONRPFPCRCCKAQCNTKQPCYLAVERCDPDLCTGGAADHMSKNVS 600  
 Db 541 NFCEKPCQCSSECONRPFPCRCCKAQCNTKQPCYLAVERCDPDLCTGGAADHMSKNVS 600  
 QY 601 CNKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISQDEADRGKVVYDK 660  
 Db 601 CNKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISQDEADRGKVVYDK 660

QY 661 YMCSEFLFNLNDPVVADTRKGNKIRFANHSVNPNCYAKVMVWNGDHRIGIFAKKAIGTGE 720  
DB 661 YMCSEFLFNLNDPVVADTRKGNKIRFANHSVNPNCYAKVMVWNGDHRIGIFAKKAIGTGE 720  
QY 721 ELFPDYRSQADALKYVGIEREMEIP 746  
DB 721 ELFPDYRSQADALKYVGIEREMEIP 746

RESULT 4  
ADN39392  
ID ADN39392 standard; protein; 746 AA.  
XX  
XX ADN39392;  
XX  
XX 17-JUN-2004 (first entry)  
XX  
XX Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:876.  
XX  
XX Human; differential expression; cancer; angiogenic disorder;  
XX fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
XX inflammatory disease; autoimmune disease;  
XX retinal neovascularization syndrome; scarring; uterine fibroid;  
XX detection; diagnosis; prognosis; drug screening; drug targeting;  
XX wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
XX vulnery; gene therapy; vaccine.  
XX  
XX Homo sapiens.  
XX OS  
XX WO2003042661-A2.  
XX  
XX 22-MAY-2003.  
XX  
XX 13-NOV-2002; 2002MO-US036810.  
XX  
XX 13-NOV-2001; 2001US-0350666P.  
XX 21-NOV-2001; 2001US-0332464P.  
XX 29-NOV-2001; 2001US-0334393P.  
XX 03-DEC-2001; 2001US-0335394P.  
XX 14-DEC-2001; 2001US-0340376P.  
XX 08-JAN-2002; 2002US-0347211P.  
XX 10-JAN-2002; 2002US-0347349P.  
XX 08-FEB-2002; 2002US-0355250P.  
XX 13-FEB-2002; 2002US-0356714P.  
XX 20-FEB-2002; 2002US-0359077P.  
XX 29-MAR-2002; 2002US-0368809P.  
XX 04-APR-2002; 2002US-0370110P.  
XX 12-APR-2002; 2002US-0372246P.  
XX 05-JUN-2002; 2002US-0386614P.  
XX 16-JUL-2002; 2002US-0396839P.  
XX 22-JUL-2002; 2002US-0397775P.  
XX 22-JUL-2002; 2002US-0397845P.  
XX 09-SEP-2002; 2002US-0409450P.  
XX  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Afar D, Aziz N, Gansburg WM, Gish KC, Glynn R, Hevezi PA;  
XX Mack DH, Murray R, Watson SR, Wilson KE, Zlotnick A;  
XX WPI; 2003-468649/44.  
XX N-PSDB; ADN39391.  
XX  
XX Determining the presence or absence of a pathological cell in a patient,  
XX useful for diagnosing, prognosing or treating cancer, comprises detecting  
XX a nucleic acid in a biological sample.  
XX  
XX Claim 12; SEQ ID NO 876; 1385BP; English.  
XX  
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
XX whose expression is upregulated or downregulated in specific cancers or  
XX other diseases such as angiogenic or fibrotic disorders, and to methods  
XX of determining the presence or absence of a pathological cell in a  
XX patient by detecting a nucleic acid at least 80% identical to those of

CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acid, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularization syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a polypeptide of the invention.  
XX  
XX Sequence 746 AA;  
SQ  
Query Match 99.7%; Score 4038; DB 7; Length 746;  
Best Local Similarity 99.7%; Pred. No. 9e-302;  
Matches 744; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MGQTGKSKGKPVCMRRKRVSEYRLQLKRPRADEVKSMFSSNRKILERTILLNQEW 60  
DB 1 MGQTGKSKGKPVCMRRKRVSEYRLQLKRPRADEVKSMFSSNRKILERTILLNQEW 60  
QY 61 KQRIQVHLITSSVSRGRSGVSDLPPTQVIFLKLTVNAVSVPIVMSPLQONF 120  
DB 61 KQRIQVHLITSSVSRGRSGVSDLPPTQVIFLKLTVNAVSVPIVMSPLQONF 120  
QY 121 MVEDEVLANIPYMGDEVLDQDGTIFIELIKNYDGKRVHGRGCFINDEIFVELVNLAQ 180  
DB 121 MVEDEVLANIPYMGDEVLDQDGTIFIELIKNYDGKRVHGRGCFINDEIFVELVNLAQ 180  
QY 181 YNDDDDDDGDDPEEREKQDLFDHDDKESRPPKFPBDDKIFRAISSMFPDGTAEEL 240  
DB 181 YNDDDDDDGDDPEEREKQDLFDHDDKESRPPKFPBDDKIFRAISSMFPDGTAEEL 240  
QY 241 KEKYEELTEQALPEECPTNIDGNPAKSVORSGLSHPTLFCRCRCFYDCEFLHPH 300  
DB 241 KEKYEELTEQALPEECPTNIDGNPAKSVORSGLSHPTLFCRCRCFYDCEFLHPH 300  
QY 301 ATPVTYRKQWETALDNKPCPGQCYQHLEGAKEPAALTAERIKTPPKRRGRRGRRLPN 360  
DB 301 ATPVTYRKQWETALDNKPCPGQCYQHLEGAKEPAALTAERIKTPPKRRGRRGRRLPN 360  
QY 361 NSRPSTPTINVLKSTQDSDREAGITGGBANDKESEKDETSSSSEANSRCQTP1KM 420  
DB 361 NSRPSTPTINVLKSTQDSDREAGITGGBANDKESEKDETSSSSEANSRCQTP1KM 420  
QY 421 KPNTEPPEVWEGSABASMPFVLIQTYDNEFCATLIGTTCQVYEFVKESSITAPA 480  
DB 421 KPNTEPPEVWEGSABASMPFVLIQTYDNEFCATLIGTTCQVYEFVKESSITAPA 480  
QY 481 PAEDVDTPPRKKRKHRLMAHCKRIQLKDGSSNHYVNTQPCDHPROPCCSFCVYIAQ 540  
DB 481 PAEDVDTPPRKKRKHRLMAHCKRIQLKDGSSNHYVNTQPCDHPROPCCSFCVYIAQ 540  
QY 541 NFCEKFCQCSSECONRPPGCRCAQCNTKQCPCTLAARECDPDLCTCGAADHDSDNVS 600  
DB 541 NFCEKFCQCSSECONRPPGCRCAQCNTKQCPCTLAARECDPDLCTCGAADHDSDNVS 600  
QY 601 CKNCISIOGSKKHLILPSPDVAGWGIPLKDPVQNGEPISEYCGEIIISODEADRGRKYDK 660  
DB 601 CKNCISIOGSKKHLILPSPDVAGWGIPLKDPVQNGEPISEYCGEIIISODEADRGRKYDK 660  
QY 661 YMCSEFLFNLNDPVVADTRKGNKIRFANHSVNPNCYAKVMVWNGDHRIGIFAKKAIGTGE 720  
DB 661 YMCSEFLFNLNDPVVADTRKGNKIRFANHSVNPNCYAKVMVWNGDHRIGIFAKKAIGTGE 720  
QY 721 ELFPDYRSQADALKYVGIEREMEIP 746  
DB 721 ELFPDYRSQADALKYVGIEREMEIP 746

RESULT 5

ADP95920  
 ID ADP95920 standard; protein; 751 AA.  
 AC ADP95920;  
 DT 07-OCT-2004 (first entry)  
 DE Human EZH2 (enhancer of zeste homologue 2), SEQ ID NO.1.  
 KM Human; EZH2; enhancer of zeste homologue 2; chromosome 7q35;  
 KM histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
 KM inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
 KM lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
 KM prophylaxis; prevention; diagnosis; cytostatic; enzyme.  
 OS Homo sapiens.  
 PN WO2004058969-A1.  
 PD 15-JUL-2004.  
 PF 22-DEC-2003; 2003WO-JP016417.  
 PR 24-DEC-2002; 2002JP-00373144.  
 PA (TAKE) TAKEDA CHEM IND LTD.  
 PI Hikichi Y, Nishizawa S;  
 DR WPI; 2004-525885/50.  
 DR N-PSDB; ADP95921, ADP95922.  
 PT Novel prophylactic and therapeutic agent of cancer, having compound or  
 PT its salt that inhibits activity of enhancer of zeste Drosophila homologue2  
 PT protein, useful for treating cancer such as lung cancer, breast cancer.  
 PS Claim 1; SEQ ID NO 1; 102pp; Japanese.  
 CC The invention relates to an agent for the prophylaxis and therapy of  
 CC cancer comprising a compound which inhibits the activity or expression of  
 CC human EZH2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
 CC EZH2 inhibitors include anti-EZH2 antibodies and EZH2 antisense nucleic  
 CC acids. The invention also relates to a kit for screening for inhibitors  
 CC of EZH2 activity or expression. EZH2 is a component of a complex with  
 CC histone methyltransferase activity which methylates histone H3 on lysine  
 CC residues at position 9 and/or 27. It has been found to be deregulated in  
 CC various tumours, and acts as a mediator of tumorigenesis, being able to  
 CC reverse repression of cyclin A which in turn mediates cell cycle  
 CC advancement. The agent of the invention can be used for preventing,  
 CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
 CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
 CC in drug screening. The agent of the invention is safe and less toxic than  
 CC prior art cancer prophylactic/therapeutic agents. The present sequence  
 CC represents human EZH2, which is referred to in the claims.  
 XX  
 XX  
 SQ Sequence 751 AA;  
 Query Match 99.5%; Score 4030.5; DB 8; Length 751;  
 Best Local Similarity 99.2%; Pred. No. 3.4e-301;  
 Matches 745; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

181 YNDDDDDGDDPBEREKEKODLEDHDDKESRPPKPPSKIFPAISMPPDKGTAEEL 240  
 DB YNDDDDDGDDPBEREKEKODLEDHDDKESRPPKPPSKIFPAISMPPDKGTAEEL 240  
 QY KKKYKELTEQQLPGALPPECTPNIDGNPAKSVQREQSILSHFTTLCRCPCFYDCFLH--- 297  
 DB KKKYKELTEQQLPGALPPECTPNIDGNPAKSVQREQSILSHFTTLCRCPCFYDCFLH 300  
 QY 298 --PFAATNTYTKRKTETALNNKPCGRCYCHLBEAKFPAALTAERIKTTPKRCGRRR 355  
 DB 301 NYSFATNTYTKRKTETALNNKPCGRCYCHLBEAKFPAALTAERIKTTPKRCGRRR 360  
 QY 356 GRLPNNSRPSTPTINVLKSDTSDRAGTETGENDKBEKEDTSSSEANSRCQ 415  
 DB 361 GRLPNNSRPSTPTINVLKSDTSDRAGTETGENDKBEKEDTSSSEANSRCQ 420  
 QY 416 TPIMKKPNIIEPPENWEGSABSMERVLIGTYDNFCALIALIGTKRGYEFKRESS 475  
 DB 421 TPIMKKPNIIEPPENWEGSABSMERVLIGTYDNFCALIALIGTKRGYEFKRESS 480  
 QY 476 IIPAPADVDTPPRKKRKRRLMAAHCRTKQLKKDSSNNVYVQPCDHPROPDSSCP 535  
 DB 481 IIPAPADVDTPPRKKRKRRLMAAHCRTKQLKKDSSNNVYVQPCDHPROPDSSCP 540  
 QY 536 CVIAQNPCEKFCQCSSECONRPGCRCAQCNTRQPCYLAVERCDPDLCTCGAADHMD 595  
 DB 541 CVIAQNPCEKFCQCSSECONRPGCRCAQCNTRQPCYLAVERCDPDLCTCGAADHMD 600  
 QY 596 SKNVSCNKSIOQSGKKHLLAPSDVAGMGIPIKDPVQKNEFISEYCGEIIISQDEADRG 655  
 DB 601 SKNVSCNKSIOQSGKKHLLAPSDVAGMGIPIKDPVQKNEFISEYCGEIIISQDEADRG 660  
 QY 656 KYTDKTCMSFLPNLNDPVVDATKGNKIRFANHSVNPCTAKVMVNGDHRIQIFAKRA 715  
 DB 661 KYTKYKMSFLPNLNDPVVDATKGNKIRFANHSVNPCTAKVMVNGDHRIQIFAKRA 720  
 QY 716 IOTGSELFPDYRSQADALATYVGIEREWEIP 746  
 DB 721 IOTGSELFPDYRSQADALATYVGIEREWEIP 751  
 RESULT 6  
 ADB65008  
 ID ADB65008 standard; protein; 707 AA.  
 AC ADB65008;  
 DT 04-DEC-2003 (first entry)  
 DE Human protein encoded by clone PUAEN20003120.  
 KM Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KM cell regeneration; membrane protein; signal transduction-related protein;  
 KM transcription-related protein; osteoporosis; neurological disease;  
 KM cancer; tumour.  
 OS Homo sapiens.  
 PN EP1308459-A2.  
 PD 07-MAY-2003.  
 PF 28-MAR-2002; 2002EP-00007401.  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehuo Y;

XX WPI: 2003-450961/43.  
DR N-PSDB; ADB63038.  
XX  
PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
PS Claim 1; Page: 222pp; English.  
XX  
CC The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or  
CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotide and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
CC the activity or expression of a protein of the encoded protein to treat diseases. The  
CC sequence presented is a protein of the invention. Note: Some of the  
CC specification data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.  
XX  
SQ Sequence 707 AA:  
Query Match 94.1%; Score 3812.5; DB 7; Length 707;  
Best Local Similarity 94.6%; Pred. No. 1.9e-284;  
Matches 706; Conservative 0; Mismatches 1; Indels 39; Gaps 1;  
QY 1 MGQTGKSEKGVCKRKVKSRYNRRLQKFRADSVKSWFSSRQKILERTETLNDW 60  
DB 1 MGQTGKSEKGVCKRKVKSRYNRRLQKFRADSVKSWFSSRQKILERTETLNDW 60  
QY 61 KORRIOPHILTSVSLRGTRCSVTSDDLPTQYIPLKTLNAVASVPIWMSPLQGNF 120  
DB 61 KORRIOPHILTSVSLRGTRB----- 82  
QY 121 MVEDETVLHNI PYMGDEVLDQDGTPIBELIKNYDGKVGHDGCGFINDEIVELNALGQ 180  
DB 83 -VEDETVLHNI PYMGDEVLDQDGTPIBELIKNYDGKVGHDGCGFINDEIVELNALGQ 141  
QY 181 YNDDDDDDGDDPEERBEKQKLEDRDDKSRPPRKPSDKIFALISMPDPDKTAEEL 240  
DB 142 YNDDDDDDGDDPEERBEKQKLEDRDDKSRPPRKPSDKIFALISMPDPDKTAEEL 201  
QY 241 KEKVELTEQOLPGALPEECTPNIDGPAKSVORQSHSFPTLPCRCGPKYDCGLHFFH 300  
DB 202 KEKVELTEQOLPGALPEECTPNIDGPAKSVORQSHSFPTLPCRCGPKYDCGLHFFH 261  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPFAALTAERIKTPPRPGRRGRGLPN 360  
DB 262 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKEPFAALTAERIKTPPRPGRRGRGLPN 321  
QY 361 NSSRSTPTINVLBSKOTDSREAGTEFGGENDVGEERBEKQKDESSSEANSRQCTPIKM 420  
DB 322 NSSRSTPTINVLBSKOTDSREAGTEFGGENDVGEERBEKQKDESSSEANSRQCTPIKM 381  
QY 421 KPNIEPPNVMESGASMPFVLTIGTYDNPFCALRLIGTKTCRQVYFRVYESSIIAPA 480  
DB 382 KPNIEPPNVMESGASMPFVLTIGTYDNPFCALRLIGTKTCRQVYFRVYESSIIAPA 441

QY 481 PAEDVDTPPKKKRKHRLMAHCRKIQLKKGSSNHVYNTQPCDHPQPCDSSCPVIAQ 540  
DB 442 PAEDVDTPPKKKRKHRLMAHCRKIQLKKGSSNHVYNTQPCDHPQPCDSSCPVIAQ 501  
QY 541 NPECKPCQSSSECONRPPGRCRAQCNTKQCPCYLAVRECDPDLCTCGAADMDSGNVS 600  
DB 502 NPECKPCQSSSECONRPPGRCRAQCNTKQCPCYLAVRECDPDLCTCGAADMDSGNVS 561  
QY 601 CRKCSIGRSKSHLLAPSDVAGWGIPIKDPVQKNEFISEYCEIISQDEADRGRKYDK 660  
DB 562 CRKCSIGRSKSHLLAPSDVAGWGIPIKDPVQKNEFISEYCEIISQDEADRGRKYDK 621  
QY 661 YMCSPFLNANDPVNATRGKTKIRPANSVNPNCYAKVMVNGDRIIGIPARALQTGE 720  
DB 622 YMCSPFLNANDPVNATRGKTKIRPANSVNPNCYAKVMVNGDRIIGIPARALQTGE 681  
QY 721 ELFPDYRYSOADALKYVIGIREMEIP 746  
DB 682 ELFPDYRYSOADALKYVIGIREMEIP 707  
RESULT 7  
ADG10764  
ID ADG10764 standard; protein: 517 AA.  
XX  
AC ADG10764;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DB Human STAT6-activating protein, SEQ ID NO:354.  
XX  
KW Human, STAT6-activating protein; drug screening; activator; inhibitor;  
KW allergic disease; inflammation; autoimmune disease; diabetes;  
KW hyperlipidaemia; cancer; infection; HIV infection;  
KW human immunodeficiency; cancer; Th1 hyperfunction; antiallergic;  
KW antiinflammatory; antidiabetic; antihypertensive; anti-infective; anti-HIV;  
KW cyostatic; gene therapy; antisense therapy; ribozyme therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200296943-A1.  
XX  
PD 05-DEC-2002.  
XX  
PF 22-MAY-2002; 2002WO-JP004949.  
XX  
PR 25-MAY-2001; 2001JP-00157043.  
PR 30-AUG-2001; 2001JP-00260681.  
PR 10-OCT-2001; 2001JP-00313175.  
XX  
PA (ASAH ) ASAH KASEI KOGYO KK.  
PI Honda G, Matsuda A, Muramatsu S, Ishizawa K,  
XX  
DR WPI: 2003-140442/13.  
DR N-PSDB; ADG10763.  
XX  
PT STAT6-activating proteins and encoded genes, applicable in diagnosis of  
PT and developing drugs to treat allergic diseases, inflammations,  
PT autoimmune diseases, diabetes, hyperlipidemia, infections e.g. HIV, and  
PT cancer.  
PS Claim 1; SEQ ID NO 354; 2080pp; Japanese.  
XX  
CC The invention relates to 242 human STAT6-activating proteins and cDNAs  
CC encoding them (ADG10411-ADG10894) and to sequences with 95% or more  
CC homology to the STAT6-activating proteins and their encoding nucleic  
CC acids. The invention also relates to recombinant vectors and host cells  
CC comprising a STAT6-activating protein-encoding nucleic acid; the  
CC recombinant production of a STAT6-activating protein; an antibody  
CC specific for a STAT6-activating protein; antisense oligonucleotides and  
CC ribozymes targeted to nucleic acids encoding a STAT6-activating protein;  
CC methods of screening for activators or inhibitors of STAT6-activating

CC proteins; drug compositions comprising a modulator of STAT6-activating  
CC protein activity or expression; and methods of treating patients by  
CC administration of the drug compositions. The STAT6-activating proteins,  
CC nucleic acids encoding them, and modulators of their activity or  
CC expression are useful in the diagnosis and treatment of allergic  
CC diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia,  
CC cancer, infections (e.g., HIV), cancer and disorders associated with Th1  
CC hyperfunction. The present sequence is related to the invention.

XX Sequence 517 AA;

Query Match 70.2%; Score 2845; DB 7; Length 517;  
Best Local Similarity 100.0%; Pred. No. 4.1e-210; Indels 0; Gaps 0;  
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 MPPDKGTAELKRYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLLPCRRRC 289  
DB 1 MPPDKGTAELKRYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLLPCRRRC 60  
QY 290 FRYDDELPHFHAHPMTYKRYKNTETALDNKPCGPOCYHLEGAKEFAALTAERIKTPPRK 349  
DB 61 FRYDDELPHFHAHPMTYKRYKNTETALDNKPCGPOCYHLEGAKEFAALTAERIKTPPRK 120  
QY 350 PGGRRGRGLPNSSRPSTPTINVLSSKOTDSREAGTETGENDKEBEKDETSSSSSE 409  
DB 121 PGGRRGRGLPNSSRPSTPTINVLSSKOTDSREAGTETGENDKEBEKDETSSSSSE 180  
QY 410 ANSRCQTPIKMKNIEPPENVMWSGABSMFVULIGTYDNCALARLIGTYTCROYEF 469  
DB 181 ANSRCQTPIKMKNIEPPENVMWSGABSMFVULIGTYDNCALARLIGTYTCROYEF 240  
QY 470 RYKESSTIAPAAEDVDTPPRRKKRKHRLMAHCKRIQLKKGSSNHYVNYPCCHPRP 529  
DB 241 RYKESSTIAPAAEDVDTPPRRKKRKHRLMAHCKRIQLKKGSSNHYVNYPCCHPRP 300  
QY 530 CDSGPCVIAONFCEKFCQSSSECONRPFPGCRCKAQCMQCPYLAVECDPDLCTG 589  
DB 301 CDSGPCVIAONFCEKFCQSSSECONRPFPGCRCKAQCMQCPYLAVECDPDLCTG 360  
QY 590 AADHWSCKNVCCKNCSIGRSGKHLALLABSDVAGWGIPIKDPVQKNEFISYCGEIIISOD 649  
DB 361 AADHWSCKNVCCKNCSIGRSGKHLALLABSDVAGWGIPIKDPVQKNEFISYCGEIIISOD 420  
QY 650 EADRRGKYDYKMGFLPTLNDPVVDATRKGNKIRFANHSYNPCYAKVMVNDHRIG 709  
DB 421 EADRRGKYDYKMGFLPTLNDPVVDATRKGNKIRFANHSYNPCYAKVMVNDHRIG 480  
QY 710 IFAKRAIQTGEELFPDYRYSQADALKYVIGIREMEIP 746  
DB 481 IFAKRAIQTGEELFPDYRYSQADALKYVIGIREMEIP 517

RESULT 8  
ADX05848  
ID ADX05848 standard; protein; 747 AA.

XX ADX05848;

DT 21-APR-2005 (first entry)

DE Cyclin-dependent kinase modulation biomarker SEQ ID NO 413.

KW cytosolic; cyclin-dependent kinase; cdk; biomarker.

OS Homo sapiens.

PN WO2005012875-A2.

PD 10-FEB-2005.

PF 29-JUL-2004; 2004WO-US024424.

XX 29-JUL-2003; 2003US-0490890P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;

DR WPI; 2005-163068/17.

DR N-PSDB; ADX05847.

PT Biomarkers useful for predicting or determining the response of a mammal  
PT to a cancer treatment comprising administration of a modulator of cyclin-  
PT dependent kinase activity.

PS Claim 5; SEQ ID NO 413; 141bp; English.

CC This invention describes a novel method of predicting or determining  
CC whether a mammal will respond or is responding to an anti-cancer agent  
CC that modulates cyclin-dependent kinase (cdk) activity. The method  
CC comprises measuring the level of one or more biomarkers selected from  
CC 2774 biomarkers given in the specification (nucleotide sequence SEQ ID  
CC NO:1246 (Genbank EST W28729) is especially preferred). The method of the  
CC invention is utilized in a kit for determining or predicting whether  
CC patient would be susceptible or resistant to treatment by an agent  
CC modulating cdk activity. The invention also describes a method for  
CC utilizing individualized genetic profiles for treating diseases and  
CC disorders based on patient's response and molecular level, specialized  
CC microarrays comprising the biomarkers described, antibodies directed  
CC against the biomarkers and a cell culture model to identify biomarkers.  
CC The cdk modulator is preferably N-5-[(5-(1,1-Dimethylethyl)-2-  
CC oxazolyl)methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-  
CC tartaric acid salt. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences. This  
CC sequence represents a biomarker used in the method of the invention.

XX Sequence 747 AA;

Query Match 64.2%; Score 2601.5; DB 9; Length 747;  
Best Local Similarity 64.6%; Pred. No. 3.8e-191; Indels 53; Gaps 10;  
Matches 490; Conservative 92; Mismatches 123; Indels 53; Gaps 10;

QY 15 WRKRVKSEYMLRLQKRRFRADDEVKSMFSSNRQKILERTTEILNOEMKORLIQPVHILTSV 74  
DB 15 WRKRVKSEYMLRLQKRRFRADDEVKSMFSSNRQKILERTTEILNOEMKORLIQPVHILTSV 74  
QY 75 GSLGTRGCVTSDL-DEPTQVILPKTNAVASPIVMSWSPLOQNFVDEBETVLANTIPY 133  
DB 75 GSHPLKCKTIBSIFPGFASQHMIMRSINTVALVIMYSWSPLOQNFVDEBETVLANTIPY 134  
QY 134 MGDEVLDQDGFIFIELINNYDGKVGHDE-----CGFINDELFEVLVNLGQYDND----- 185  
DB 135 MGDEVKEDETFIIBELINNYDGKVGHDEMIIGSVLISDAVFLBLVDNLNIGYSDBEERGH 194  
QY 186 -----DDDDGDDPEREERKOKLDEHDDKESRPRKPFSPDKIFPAISMPFDKGT 237  
DB 195 NDTSDGKQDDSKDLPLVTRKKRRAIAEGNK--KSK--KQPNMIMIFALISMPPENGV 250  
QY 238 ELKRYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLLPCRRRCFYKDCFLH 297  
DB 251 DDMKERYELTEMSDPMNLPPQCTPNIDGNPAKSVQREQSLSHFTLLPCRRRCFYKDCFLH 310  
QY 298 PFHATPTNYKRYKNTETALDNKPCGPOCYHLEGAKEFAALTAERIKTPPRRPGRRRGR 357  
DB 311 PFHATPTNYKRYKNTETALDNKPCGPOCYHLEGAKEFAALTAERIKTPPRRPGRRRGR 364  
QY 358 LPNNSRPSTPTIN-VLESKOTDSREAGTETGENDKEBEKDETSSSSSEANSRCOT 416  
DB 365 HHIVSASGNSASAAVALETGSDSDRDTG-----NDMASSSEANSRCOT 409  
QY 417 PIKKNPNIEPP-----ENVMWSGABSMFVULIGTYDNCALARLIGTYTCROY 467  
DB 410 PTKOKASAPAPQOLCVNEAPSEPVEMTGAEBSLFVFFGTYFPNFCISLARLIGTYTCROY 469  
QY 468 EFRVYKESSTIAPAAEDVDTPPRRKKRKHRLMAHCKRIQLKKGSSNHYVNYPCCHPR 527



```

Db 195 NDTSDGKDDSDKEDPVRKRKRHAIBGNK--KSKK--KQFNDMIFSAISMFPENGV 250
Qy 238 BELKEKYELTEQOLPGALPPECTPNIDGNPAKVSQVREOSLSHFTLLFCRCFCYDCLH 297
Db 251 DDMKERYELTEMSDPNALPQCTPNIDGNPAKVSQVREOSLSHFTLLFCRCFCYDCLH 310
Qy 258 PPHATPNYKRRNTTALDNKPCGQCOHLEGAPFAALTAERIKTPPKRPGRRGR 357
Db 311 PPHATPNYKRRNKIKTEPEPCGIDCFLLEGAEYA-----MLNHPRSKSGRRRRR 364
Qy 358 LPNNSRSTPTIN-VLESKTDSDRBAETETGGENNDKEBEKDETSSESEANSRCOT 416
Db 365 HHIVASCSNASASAVAETKEGDSRDYD-----NDMASSSEANSRCOT 409
Qy 417 PIKMKNTIEPP-----ENVENSGABAMFRLIGTYDNPFCALRLIGTKTCROY 467
Db 410 PTKQKASAPAPOLCVVEAPSEFVETWGAESLFRVHGTYFNNPFCISARLIGTKTCROY 469
Qy 468 EFRVSESIAPAPAEVDTPPRKKGKGRLLMAACRKLQLKQDSSNHVNYNYQPCDHP 527
Db 470 QPAAVE-SLILKLPDELMYPSQKKRGRLLMAACRKLQLKQDSSSTQVYNYQPCDHP 528
Qy 528 QPCDSSCPVLAQNFCEKRCQCSQNRPGCRCKAQCNWKQPCYLAVERCDPDLCLT 587
Db 529 RPCDSTCFIMQNFCEKRCQCNPD-----LNECDPDLCLT 564
Qy 588 CGAADHMDSKNVSCKNCSIQRSKSHLLAPSDVAGWGIPIKDPOKNEFISEYCGELIS 647
Db 565 CGASHMDCKVVSCKNCSIQGLKSHLLAPSDVAGWGTFIKESVQKNEFISEYCGELIS 624
Qy 648 QDEADRRKGVYDKTKCSFLFNLANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHR 707
Db 625 QDEADRRKGVYDKTKCSFLFNLANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHR 684
Qy 708 IGIPAKRAIQTGESELPFDYRYSQADALKYVGIEREMEI 745
Db 665 IGIPAKRAIQAGESELPFDYRYSQADALKYVGIERETDV 722

RESULT 10
ABB62427
ID ABB62427 standard: protein; 760 AA.
XX ABB62427;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 14073.
DE
XX Drosophila melanogaster polypeptide SEQ ID NO 14073.
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL06530.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX

```

```

PT interactions.
XX
XX Diaclosure; SEQ ID NO 14073; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
XX sequences (AB16176-AB16175) and the encoded proteins (AB57737-
XX AB577072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WPI at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 760 AA;
XX
XX Query Match 53.2%; Score 2156.5; DB 4; Length 760;
XX Best Local Similarity 53.9%; Pred. No. 7.3e-157;
XX Matches 432; Conservative 86; Mismatches 176; Indels 107; Gaps 14;
XX
Qy 8 SEKGPVCKRRKVKSEYNRLQKFRRADEVKSNFSSNRQKILERTETL--NDEWKOR 64
Db 3 STKVPEWKRRKVKSEYIKIRQKKYKRADEIKAMIRWDEHNHNVQDLYCESKVMQAKP 62
Qy 65 IQPVHILTSVSLNGTRCSTSDLPPT--QVPLKLANAVASVPIMYSPLQONFMV 122
Db 63 YDPFHV-----DCVKRAEVTSYNGIPSGQKVPICVINAVTPIPTMYTAAPQONFMV 115
Qy 123 EDETVLNIPYMGDEVLDDQGTFLBELIKYNDGVHGDREGCFINDELFVELNAL----- 178
Db 116 EDETVLNIPYMGDEVLDDQGTFLBELIKYNDGVHGDREGCFINDELFVELNALMSY 175
Qy 179 -----QYNDP--DDDDGDDP-----EEREE 198
Db 176 SKELBEAPGTAIKTETLAKSKQGEDDGVVDADQESPMKLEKTDKSGDLTEVEKE 235
Qy 199 KQKLEHRRD-----KESRPRKPPSDKIFBAISSMPDGTBELKEKTELTOOLP 253
Db 236 TEBPLETEDADVKEVBEVQKLPFPAPIIFQALISANFPDGTQELKEKTELTEHDP 295
Qy 254 GALPPECTPNIDGNPAKVSQVREOSLSHFTLLFCRCFCYDCLHPF--HATPNYKRRNT 311
Db 296 -ERPQCTPNIDGIKASVSHERKTHMSHFTLLFCRCFCYDCLHPHGHAPNIOKRYR 354
Qy 312 ETALDNKPCGQCOHLEGAPFAALTAERIKTPPKRPGRRGRRLPNNSRSTPTIN 371
Db 355 ELKFPAPCSNSCYMLIDGMKEXLAADS--KTP----- 386
Qy 372 VLESKTDSDRBAETETGGENNDK-----EEREKDETSSESEANSRCOTPIKMKNI 424
Db 387 -----IDSCHASSESDNSNSQPSNKFENHENSNGKGLVNSAAVAEISINAGWNI 440
Qy 425 EPPENVMSCGAEMFVLIGTYDNPFCALRLIGTKTCROYFVRESIIAPAPAD 484
Db 441 TSTQCV-WTGADQALVYVHLKRVYKANYCALAHNMLTKTCROYVFAQKEDAEFEDLRQ 499
Qy 485 VDTPRKKKKRGRLLMAACRKLQLKQDSSNHVNYNYQPCDHPROPSCPVLAQNFCE 544
Db 500 DFTPRKKKKKQRLMSLHCRKLQLKQDSSNHVNYNYPCDHPGHGCMONGSCIQTNFCE 559
Qy 545 KFCQSSSCQNRPGCRCKAQCNWKQPCYLAVERCDPDLCTGAAHDMDKNVSCNVC 604
Db 560 KFCNCSDDQNRPGCRCKAQCNWKQPCYLAVERCDPDLCTGAAHDMDKNVSCNVC 618
Qy 605 SIORGSKHLLAPSDVAGWGIPIKDPOKNEFISEYCGELISQDEADRRKGVYDKTKCS 664
Db 619 CVQKGLHLLMAPSDVAGWGIPIKESQKNEFISEYCGELISQDEADRRKGVYDKTKCS 678
Qy 665 FLFNLANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIPAKRAIQTGESELP 724
Db 679 FLFNLANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIPAKRAIQTGESELP 736

```







ADY25049  
ID ADY25049 standard; protein; 880 AA.  
XX  
AC ADY25049;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polypeptide seqid 72833.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.  
XX  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
PD 19-FEB-2004.  
XX  
PF 28-APR-2003; 2003US-00425114.  
XX  
PR 06-MAY-1999; 99US-00304517.  
XX  
PR 05-NOV-2001; 2001US-00985678.  
XX  
PA (LITUJ/) LIU J.  
XX  
PA (ZHOU/) ZHOU Y.  
XX  
PA (KOVA/) KOVALIC D. K.  
XX  
PA (SCRE/) SCREEN S. E.  
XX  
PA (TABA/) TABASKA J. E.  
XX  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
XX  
DR WPI, 2004-180133/17.  
XX  
PT New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
PS Claim 1; SEQ ID NO 72833; 15pp; English.  
XX  
CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp://seqdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 880 AA;  
Query Match 20.5%; Score 828.5; DB 8; length 880;  
Best Local Similarity 28.8%; Pred. No. 1.7e-54;  
Matches 251; Conservative 113; Mismatches 267; Indels 241; Gaps 38;  
OY 25 RLRLQKRRRADEVKSM---FSSNRQKILTEILINQGMKORRIGPVHILTVSSLSRGTR 81

DB 59 RLSELKRLQGERRVSISKEFENRKY-----DAHVS-----FSSAASRATA 103  
OY 82 ECSVTSDDLPPTQVYPLKTNAAVASVPIYMSWSPLOQNFVDETVLNI-----P 132  
DB 104 EDNGNSNMLSSRMMPFLCKLNG-----FSGVGDSDRYVPTKVIVISASVYLPIAERLIP 156  
OY 133 YMGDEVLDQDGTFLIELLNKVDGKVGHDREGFINDLFEVLVNLVAGQYND--DDDG 190  
DB 157 YTTWIFLDRNQPMAB-----DQSVGR-----QLYE-----OHGSETLICDSE 197  
OY 191 DDPEREKQKQLEDHRDDEKSRPRKPSDKIFEALISMPDGTAEELK-----241  
DB 198 EEPEREKRRFSEB-----ESITIMLQGEY-----GMBEYVDALCQLLSV 240  
OY 242 -----EKYKELTEQQLPGALPPECTPNIDGPNK-SVQRQ-----SIHSTFLPCRCF 290  
DB 241 DASDILERYNELKLD-----KONTSEFSNGSFGKLSLEKGLGALDSFDNLFCRRCL 294  
OY 291 KYDCELHPHATPMTYRKXN-----TETALDNPCGPGCQYHLEAKE-----PAA-----336  
DB 295 VFDRLHGCQPLISASEKQPYMSDYEGDRPCSKGCTYLQKAVREVPETCSNPFASKAE 354  
OY 337 -----ALTAER-----IKTPPKRPG-----GRR-----355  
DB 355 KASEBECKAVSDVPHAAAGVSLQVEKTDIGIKNDSSSGVQDHEGIRKREVPILKD 414  
OY 356 -GRLPNNSSRP-----STFTINVLBSKQDSDREAGTGTGGENNDEEEK 400  
DB 415 SNDLPNLNSNKKQKTAASPTKMSFVNSVPSLD--QALDSTKGDQGGTDTNKNRDSADAK 472  
OY 401 K-----DETSSSSSANSRCQTP-----IKMKRIEPPENVEWGSAGARSFVILG 445  
DB 473 EVGEPPIPNVSVDGSSSICQPHHSGNGALIIAAMSETSRP--STEMPIEKDLYLKVE 530  
OY 446 TYDNFCAIAR--LIGTCRQYEF-RVKESSIIAPA-----PAEDVD--TP 488  
DB 531 IFRNSCLIANNLGLKTCILDVSNMYRKNVSVFRSSTNNLLDDORTPGDNDSDVP 590  
OY 489 PR---KKGRKRIW-----AAHCKRIQLKDGSSNHVINYOPCDHPRPCDSSCPVY 538  
DB 591 PRTLPRFRKGTIRKLTSTKSAHPVSWKTAGKNOSCQKQYPCG--CLSMCGKDCPCLT 649  
OY 539 AQNFCERKCCSSSCNRRPFCRC--KAOCNKKQCPCTIAYECCPDLCTLGCADHMSK 597  
DB 650 NETCEKRCGCSKCKNRFRCGCAKSCQCSRCQCPAAGCECPDVCRNC-----W---701  
OY 598 NVSCNKSISI---ORG-----SKKHLIAPSDVAGMGIFFIKDPYQKNFISBY 641  
DB 702 -VSCGDGSLGAPFRRGQCGMMLLRQQRILIGKSDVAGMGAFLKNSVSKREYLGEY 760  
OY 642 CGEIIISQDEADRRGKYDKTWCSFLFNILNDDVVDATRKCKIRFANHVNPNCYAKVM 701  
DB 761 TGEIISHHEADKRRKTYDRANSSFLPDLNDQYVDAQRKGBKLKFAHNSAKPNCYAKMF 820  
OY 702 VNGDHRIGIRPAKALQGEELFFPYRYSQADA 733  
DB 821 VAGDHRGIVANERIEASEBELFYRYRGPDQA 852  
RESULT 13  
AA014530  
ID AA014530 standard; protein; 893 AA.  
XX  
AC AA014530;  
XX  
DT 16-MAY-2002 (first entry)  
XX  
DE Protein of Nez2 (Maize E(z)-like 2).  
XX  
KW Maize; plant; genetic engineering; polycomb group; Pcg; Maize E(z)-like;  
KW transgenic; antibody; marker; crop improvement; agronomic trait;  
KW homeotic gene expression; immunogen; screening; affinity chromatography;

KM Mez 2; chromosome 9 (bin9.04).  
XX  
OS Zea mays.  
XX  
FH Key Location/Qualifiers  
FT Domain 142..254  
FT Domain /note= "EZD1 domain"  
FT Domain 267..310  
FT Domain /note= "EZD2 domain"  
FT Domain 535..571  
FT Domain /note= "SANT domain"  
FT Region 608..636  
FT Region /note= "NLS region"  
FT Region 672..724  
FT Region /note= "Cys-rich region"  
FT Domain 745..870  
FT /note= "SET domain"  
XX  
PN MO200206322-A2.  
XX  
PD 24-JAN-2002.  
XX  
PR 16-JUL-2001; 2001WO-US022713.  
XX  
PR 17-JUL-2000; 2000US-0218745P.  
XX  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA (PION-) PIONEER HI-BRED INT INC.  
PA (MTNU ) UNIV MINNESOTA.  
XX  
PI Kaepler SM, Springer NM, Helentjars TG, Phillips RL;  
XX  
XX WPI; 2002-216997/27.  
DR N-PSDB; AAK98824.  
XX  
FT Novel polycomb group nucleic acids and polypeptides from maize useful for  
FT producing transgenic plants and to repress expression or prevent the  
FT expression of a desired target gene in specific tissue in plants in vivo.  
XX  
PS Claim 8; Fig 2B; 80pp; English.  
XX  
XX The invention relates to plant genetic engineering. More specifically,  
CC the invention relates to isolated and purified polycomb group (pcg) genes  
CC from Zea mays (maize), termed Mez1 and Mez2 (Maize R(2)-like 1 and 2)  
CC genes comprising a sequence of 3180 and 3030 bp respectively, defined in  
CC the specification, or its conservatively modified and polymorphic  
CC variants. The pcg genes of the invention are useful in constructing  
CC recombinant expression cassettes which can be introduced into a desired  
CC plant, to produce a transgenic plant. Promoters derived from the Mez1 or  
CC Mez2 genes are useful to direct expression of heterologous sequences. The  
CC modified variants of the pcg genes are useful for generating or selecting  
CC antibodies immunoreactive to the non-variant polypeptide and to follow  
CC the segregation of chromosome regions and in marker assisted selection  
CC methods for crop improvement. Mez1 and Mez2 proteins can be used to  
CC repress the expression or prevent the expression of a desired target gene  
CC in specific tissue in a plant in vivo, to regulate homeotic gene  
CC expression in plants to create novel plants having improved agronomic  
CC traits. Protein derived from a pcg gene of the invention is useful as an  
CC immunogen for producing monoclonal or polyclonal antibodies which are  
CC useful for screening plants for the expression of a Mez1 or Mez2  
CC polypeptide, and for affinity chromatography for isolating a Mez1 or Mez2  
CC polypeptide. This sequence represents the protein of Mez 2 of the  
CC invention  
XX  
XX  
SQ Sequence 893 AA;  
  
Query Match 20.0%; Score 810; DB 5; Length 893;  
Best Local Similarity 26.8%; Pred. No. 4.5e-53;  
Matches 260; Conservative 108; Mismatches 249; Indels 352; Gaps 42;  
  
QY 3 QTKGKSEKG-----PYCKRKVKSEYMKLRQLKRPRAAEVSKMPSNOKLLE 51  
DB 12 QRSKRSDOGMGKDAAAASVPI-----HANTLTL-----IKOVOSGLAYIKE 54

QY 52 RTBILNQEMKQRR---IQPVHILTSVSLNGT-----RECSVTSDL----- 89  
DB 55 KLEAV-NRKTLOHSHSCSLFDVAAAAEVAS -RGTDGNGALSGRAABRQCG--SDLANGIGER 110  
QY 90 -----DPTQVITPLKTLNAAV-----SVPIMVSWSPLOON-FMVVEDT 126  
DB 111 DVVSVHEENLATGTALSSSGATQRTIVFVKLPVLEKIPTYTWTFLDKNQRMADQS 170  
QY 127 VL--HNIPY--MGDEVL--DQDGFIEE-----LIKNDGKV-----HGRREGCFIN 167  
DB 171 VGGRRRIYDTGVNGEALICSDSDDEIPEPEBEKHFFTKSGDHLIMRATQDHG-----LN 224  
QY 168 DEIFVELVNALGQYNDDDDDDDDDPEREREKQDLERHDDKESRPPKFP--DKI 223  
DB 225 QEV---VNYLCQF-----IGATPSHIEERSEVLPE-KNEKISGSSDKIESRLSDKT 272  
QY 224 PEAISSMFPDQKGTAELEKKEKLTBQQLFGALBPPECTPNIIDPNNAKSVQREGSLSHFHT 283  
DB 273 MDAAV-----LDSFDN 282  
QY 284 LFCRCRCFYDCFLH-----PFATPTYRKKTETALDNKPCGPQCY---Q 326  
DB 283 LFCRCCLVPCDRHLGCSQNLVFPCEKQPYSPDPENK-----KPCGHLCTYAFPO 332  
QY 327 HLEGAKS----- 333  
DB 333 WRBGFKEHDDGLAGATYTWESGTAQVYDVVWVYESDSNRQGNIRSWTLVTSGPK 392  
QY 334 FAALTLTAERIKTP-----KPPGGRRRGLPNNSSRPSTPTI 370  
DB 393 IISVSABESTTTPADISETENVSSDLPPSGLRKHKISKGPFRHHSFGKQKVTFTSDI 452  
QY 371 -----NVLESKDTSDREAGTETGEGENNDKEEEKKDETSSSSSEANSRC-QTPIK 419  
DB 453 SFEGSIANKLSIFIRDT---RLESRESG-----DLKRIIDESTKTKSRKMDGSEBPAT 504  
QY 420 MKENIPEPPN-----VE-----WSGAESMFRVLIGTYDNFCALAR--LIGTYTCR 464  
DB 505 TMENVGRQSNKRVYSTKNFLBSTLSGWSALERDYLKGIETIFGKNSCLIANRLLSGLKTCI 564  
QY 465 QVYFRFKESIIAP-----ABDVPTPPKKRKR-----LM- 499  
DB 565 EVANYMTNNGAMAKRPULANKSISGDPAENBODYEMQMAARTIYRRGRNRLKXTWK 624  
QY 500 -AAHCRKIQLKQDSSNHNVNYQPCDHPROPCDSSCPVLAQNECEKRCQSSSECONRFP 558  
DB 625 SAGH-PTVRKRTDGGKQCTGYQYSPC-ACQGMCGKDCQADKGTCEKXCGCSKSKXKFR 682  
QY 559 GCRC-KAQCTKQCPCYLAVERCDPDLCTTCGAADHDS-----KXVSCNKC 604  
DB 683 GCHCAKQCRSRQPCFPASRECDPVCRCN-----WVSCGDGSLGEPRLARGDGYQCGNM 737  
QY 605 SIQGSKKHLLAASDVAGWGIPTKDPVQKNEFISEYOGHIIISODBADRQKVDKYMCS 664  
DB 738 KLLKQOQRILLGSDVAGWAFIKNPVKNQDVIGETTGELISKEKDKRKLTIDRANSS 797  
QY 665 FLPLINLDFVDATRKNGKIRFANHSVNPYATKVMVNGDHRIGIPAKRAIQTGBELFF 724  
DB 798 FLPLNDQGYVDARDKDKLKFANHSSNPNTYAVMLVAGDHRIGIYAKHEIIBASELEFY 857  
QY 725 DYRYSGQADA 733  
DB 858 DYRYGPDQA 866  
  
RESULT 14  
ADM48181  
ID ADM48181 standard; protein; 891 AA.  
XX  
XX ADM48181;  
DT 03-JUN-2004 (first entry)

XX Polypeptide sequence #231 useful in producing transgenic plants.  
 XX  
 DB Plant; maize; rice; sorghum; thale cress; soybean; wheat; transgenic;  
 KW  
 KW osmotic stress; sugar transport; cell cycle pathway; plant height;  
 KW carbohydrate transport; crop productivity; plant growth;  
 KW stress resistance; disease resistance; insect resistance; heat tolerance;  
 KW nitrogen assimilation; water stress tolerance;  
 KW photosynthetic carbon fixation; virus resistance; gene therapy.  
 XX  
 OS Zea mays.  
 XX  
 PN US2003233670-A1.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 04-DEC-2002; 2002US-00310154.  
 XX  
 PR 04-DEC-2001; 2001US-0337358P.  
 XX  
 PA (EDGE/) EDGERTON M. D.  
 PA (CHOM/) CHOMET P. S.  
 PA (LACC/) LACCETTI L. B.  
 XX  
 PI Edgerton MD, Chomet PS, Laccetti LB;  
 XX  
 DR MPI: 2004-061374/06.  
 DR N-PSDB; ADM47813.  
 XX  
 PT New polynucleotide, useful for manipulating plant protein quality,  
 PT improving plant growth, yield and crop productivity or grain composition  
 PT or producing plants with improved properties.  
 XX  
 PS Claim 8; SEQ ID NO 599; 144bp; English.  
 XX  
 CC The present invention relates to polynucleotide sequences, and the  
 CC proteins they encode. The sequences are isolated from a variety of  
 CC organisms such as plants (e.g. maize, rice, sorghum, thale cress,  
 CC soybeans and wheat), cyanobacteria, bacteria, yeast and other fungi. The  
 CC polynucleotide and polypeptide sequences of the invention are useful in  
 CC the production of transgenic plants that have improved properties. Also  
 CC disclosed are methods of producing fertile transgenic plants, preferably  
 CC maize, with desired phenotypes. The polynucleotide and polypeptide  
 CC sequences are useful for improving plants by providing protection against  
 CC osmotic stress, improving altering sugar transport and/or metabolism,  
 CC modifying the cell cycle pathway, reducing plant height, modifying  
 CC carbohydrate transport, improving crop productivity, improving plant  
 CC growth and stress resistance, improving disease resistance, improving  
 CC insect resistance, improving cold or heat tolerance, improving nitrogen  
 CC assimilation, improving stalk strength, improving water stress tolerance,  
 CC improving photosynthetic carbon fixation, improving biotic and abiotic  
 CC stress resistance, improving resistance to oxidative stress, providing  
 CC increased vigor, reducing senescence, and conferring virus resistance.  
 CC The present sequence represents a polypeptide sequence of the invention.  
 CC Note: The sequence data for this patent is not provided in the printed  
 CC specification but is obtained in electronic format from the USPTO website  
 CC at seqdata.uspto.gov.  
 CC  
 XX  
 SQ Sequence 891 AA;  
 XX  
 Query Match 19.9%; Score 805; DB 8; Length 891;  
 Best Local Similarity 26.8%; Pred. No. 1,1e-52;  
 Matches 259; Conservative 108; Mismatches 250; Indels 350; Gaps 42;  
 XX  
 QY 3 QTGKSKRG-----PVCWRKRVKSEYMLRLDKRRFRADKVMSSNRQKILE 51  
 DB 12 QRSKRSQGMGKDAASVPI-----HANLTQ-----IQVSGRLAYIK 54  
 QY 52 RTEINQEMKOR---IQPVHILTSVSLRG-----RCGVTSDL----- 89  
 DB 55 KLEV-NRRTQNHSCSLFDVAAAEVAS-RGTDGNNALSGRAAEQCG--SDLANGIGER 110  
 QY 90 -----DFPTQVILPLKTLNANA-----SVPIWMSBPLQGN-FWVVEDET 126

DB 111 DVVSVQENLANTGLATLSSGATAQTIVRFVKPLVLEKIPYTTWIFLDKNGRADDOS 170  
 QY 127 VL--HNIPY--MGDEVL---DODGTFIE-----LKNYDGK-----HEDREGFTN 167  
 DB 171 VGRRRRIYDITVGNALICSDSDEIPEPEBEKHFFPTGEBDLIRATQDHG-----LN 224  
 QY 168 DEIFELVNAAGVYNDDDDDGDDPEBEBEKQXLEHHRDDKESRPRKPPS-----DKI 223  
 DB 225 QEV---VNVLCQF-----IGATPSEIBKSEVLF-KNEKHSSSDKIESRLSLDKT 272  
 QY 224 FEALISWEPDKGTAELEKYEKELTEQQLPGALPECTPIIDGPAKSVQSEQSLSFPT 283  
 DB 273 MDV-----IDSTDN 282  
 QY 284 LFCRRCFKYDCLH-----PFAHTPTYYKKNMTETALDNKPCPGQY-QHLE 329  
 DB 283 LFCRRCLVFDCLHGCSONLVFPCBKQYSPDPDNK-----KPCGHLCYLRWRE 332  
 QY 330 GAK-----FAA 336  
 DB 333 GFKEMHDDGLAGATTWESGTASQVVDVNVYSESDSNRQKINIRSMTLVGTSGSKII 392  
 QY 337 ALTAERIKTPP-----KRCGRRRGRLPNNSRPSTPTI-- 370  
 DB 393 SVSABESTTTBSADTSETENVSDDLPPSLAKHKIKSKGPRYREHSPKGRQVFTSDISF 452  
 QY 371 -----NVLESKOTDSREAGTETGGENNDKEBEKKEDTSSSSSEANSRC-QTPIRK 421  
 DB 453 EGNIMNKLSIPIRDT--RLESRESG-----DLRILIDSTYKTSRKDMGESPATM 504  
 QY 422 PNIPEPEN-----VSGAASMPRVLIQTYNFCALAR--LIGTTCQY 466  
 DB 505 ENVRGQSKVSTKNFLESTLSCWALERDLYLKGIELFGKNSCLIANLNLGLTKCLEV 564  
 QY 467 YEFRVKESSIIAPAP-----AEDVTPPRKKRKR-----LW--A 500  
 DB 555 ANYMNNPMAAKRPLANKSISGPAHEBQYMDMAAKRIYRRGRNKLKTYMWSA 624  
 QY 501 AHCRKIQLKQDSSNHYVNYQPCDHPQPCDSCPCVIAQNFCXKPCQSSSECONRPFQC 560  
 DB 625 GH-FVRRRTDQKQCYQYSPC-AQQMCGKDCADKGCCEKCYCGCSKSKNKKFRGC 682  
 QY 561 RC-RAQCTTKCCPCVLANREDDPLCLTCGAADHMS-----KNVSKNCSI 606  
 DB 683 HCASQCSKRCPCFASRBCDPVCRNC-----WVSGDGLSEPLARGDGYQGNKL 737  
 QY 607 QRSKSHLLAPSDVAGGIFIKDPVQKNBFISEYCGEIIISODEADRRGKYVDKYMCSFL 666  
 DB 738 LKQOQRIILGRSDVAGGAFIKPVPKNDYLGRTGELSHKPADRKGIYDRANSFL 797  
 QY 667 FNLANPVDATRKGNKIRFANHSVNPCKAVKVMVNGDHRIGIIPAKAIGTEBELFPDY 726  
 DB 798 FDLNDQYLDAYRRGDKLKFANHSNPPCKAVKWLVA GDHVGIVYAKHIIASELFPDY 857  
 QY 727 RYSGADA 733  
 DB 858 RYGPDA 864  
 XX  
 RESULT 15  
 ADTS7263  
 ID ADTS7263 standard; protein; 931 AA.  
 XX  
 ADTS7263;  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE Plant polypeptide, SEQ ID 7340.  
 XX  
 KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
 KW disease resistance; galactomannan production; plant growth regulator;  
 KW heat tolerance; herbicide tolerance; lignin production;



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:21:54 ; Search time 29.427 Seconds  
(without alignments)  
2439.179 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGQTGKSEKGPVCKRKRKRVK.....RYSQADALKYVGIEREMRIP 746

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the query being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4038	99.7	746	2	G02838
2	828.5	20.5	856	2	T52415
3	813.5	20.1	898	2	T01503
4	795.5	19.6	902	2	T01127
5	726	17.9	689	2	T52060
6	618	15.3	775	2	T21436
7	306.5	7.6	2144	2	S71490
8	300	7.4	3968	2	A44265
9	298	7.4	3869	2	A48205
10	297	7.3	920	2	T41282
11	291.5	7.2	1080	2	S48961
12	287.5	7.1	1802	2	H88444
13	280	6.9	528	2	E36795
14	279	6.9	1767	2	T00458
15	275.5	6.8	352	2	T47820
16	275.5	6.8	739	2	A88445
17	272.5	6.7	344	2	T00695
18	271	6.7	2588	2	T14342
19	265.5	6.6	477	2	C85361
20	261	6.4	408	2	T16601
21	254	6.3	1001	2	S30385
22	248.5	6.1	1193	2	A86193
23	247.5	6.1	3828	2	T13857
24	247	6.1	3759	2	A35085
25	245.5	6.1	798	2	T38490
26	244	6.0	635	2	S47004
27	244	6.0	733	2	S56951
28	238.5	5.9	2422	2	T12687
29	238	5.9	186	2	B84723

30	230.5	5.7	788	2	C84616
31	226.5	5.6	4957	2	T03455
32	226.5	5.6	5262	2	T03454
33	224.5	5.5	203	2	T02416
34	224	5.5	1590	2	B87754
35	218.5	5.4	490	2	T43700
36	214.5	5.3	490	2	T43745
37	209.5	5.2	794	2	D84765
38	202.5	5.0	898	2	T26577
39	198	4.9	669	2	P96756
40	198	4.9	954	2	G86312
41	196	4.8	401	2	T32348
42	189.5	4.7	1312	2	T33758
43	187	4.6	2561	2	T24864
44	174	4.3	2434	2	S44861
45	171.5	4.2	429	2	C84640

similar to mammal  
ALR protein - huma  
ALR protein - huma  
probable SRT-domai  
protein C43B11.3 l  
mating-type loci a  
clra protein - fis  
similar to mammal  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
DNA topoisomerase  
similar to mammal

ALIGNMENTS

QY	1	MGQTGKSEKGPVCKRKRKRVKSYMLRQLKRRRADVYKSMFSSNROKILERTILNQEW 60	99.7% ; Score 4038 ; DB 2 ; Length 746 ;
DB	1	MGQTGKSEKGPVCKRKRKRVKSYMLRQLKRRRADVYKSMFSSNROKILERTILNQEW 60	Best Local Similarity 99.7% ; Pred. No. 4.3e-237 ;
QY	61	KORRIQPHILITVSLSIGTRCSTSDLDPPYIPIKTLNVAVASVPIWMSPLQONF 120	Matches 744 ; Conservative 0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;
DB	61	KORRIQPHILITVSLSIGTRCSTSDLDPPYIPIKTLNVAVASVPIWMSPLQONF 120	
QY	121	MVEDETVLNTPYMDDEVLDODGTPIEBILIKNYDKVAGDRCEGINDBEIFELVNALGQ 180	
DB	121	MVEDETVLNTPYMDDEVLDODGTPIEBILIKNYDKVAGDRCEGINDBEIFELVNALGQ 180	
QY	181	YNDDDDDDDGDDPEERBEKQKLEBDHDDKESRPPKPSDKIFPAISMPFDDKGTABEL 240	
DB	181	YNDDDDDDDGDDPEERBEKQKLEBDHDDKESRPPKPSDKIFPAISMPFDDKGTABEL 240	
QY	241	KEKYKELTEQQLPGLAPRECPINIDGNPAKSVQRQSLHSFTTLFCRCFCFKDCLLHPH 300	
DB	241	KEKYKELTEQQLPGLAPRECPINIDGNPAKSVQRQSLHSFTTLFCRCFCFKDCLLHPH 300	
QY	301	ATPNTYKAKNTETALDNKPCGPOCYQHLEGAKEFAAALTABRIKTPPRPGRRGRGLPN 360	
DB	301	ATPNTYKAKNTETALDNKPCGPOCYQHLEGAKEFAAALTABRIKTPPRPGRRGRGLPN 360	
QY	361	NSSRPSTPTINVLBSKOTSDREACTETGGENNDKEEKKOETSSSSSEANSRCOTPIK 420	
DB	361	NSSRPSTPTINVLBSKOTSDREACTETGGENNDKEEKKOETSSSSSEANSRCOTPIK 420	
QY	421	KPNIEPPNENMSGABAMFVILGTYDNCALARLIGTICRQVYEFKRVESSTIAPA 480	

```

Db      421  |KNIIEPPENVEWSGAEASMFRLVIGIYDNFCAIANLIGTKICRQYIERRVKESSTIIAAR 480
QY      481  |PAEDVDTPPRKKKKRKKRLMAAHCRLKIQLKKDSSNNHVINYQPCDHBPQPDCSSPCVIAQ 540
Db      481  |PAEDVDTPPRKKKKRKKRLMAAHCRLKIQLKKDGSNNHVINYQPCDHBPQPCDSSPCVIAQ 540
QY      541  |NCEKYPCCSSSCQONPFCGCRCAQCNQTCQPCYLAVRECDPDLCTCGAADMDSKSNYS 600
Db      541  |NCEKFCQCCSSBCQNRPFCCRCCKAQCNQTCQPCYLAVRECDPDLCTCGAADMDSKSNYS 600
QY      601  |CNKCSIQRGSKKHLIAPSDVAWGMIPIQDPVQKNEFISYCSGIIISQDADPRGRKYTDK 660
Db      601  |CNKCSIQRGSKKHLIAPSDVAWGMIPIQDPVQKNEFISEYCSGIIISQDADRGGRKYTK 660
QY      661  |YWCSEFLFNLANNDPVVDATRKGNKIRIPANHSVNNCYCAKMMVNVGDHRIGIFAPRAALQTGE 720
Db      661  |YWCSEFLFNLANNDPVVDATRKGNKIRIPANHSVNNCYCAKMMVNVGDHRIGIFAPRAALQTGE 720
QY      721  |ELFPDPRYSQADALAKVGIEREKEIP 746
Db      721  |ELFPDPRYSQADALAKVGIEREKEIP 746

RESULT 2
752415
polycomb protein EZAI [imported] - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 24-Oct-2000 #sequence_rev1sion 24-Oct-2000 #ext_change 09-Jul-2004
C|Accession: T52415
R:Blitodeau, P.; Luo, W.; Dennis, E.S.; Peacock, W.J.; Chaudhury, A.M.
submitted to the EMBL Data Library, October 1998

```

A,Accession: T52415  
A,Reference number: 226069  
A,Status: preliminary; translated from GB/EMBL/DBJ  
A,Molecule type: mRNA  
A,Residues: 1-856 <BIL>  
A,Cross-references: UNIPROT:O9ZSN8, UNIPARC:UPI00000A5399, EMBL:AF100163, PIDD:PAD09108

Query Match	20.5%;	Score 828.5;	DB 2;	Length 856;
Best Local Similarity	28.8%;	Pred. No. 1.3e-42;		
Matches 251; Conservative	113;	Mismatches 267;	Indels 241;	Gaps 38;

[illegible]

```

Db 391 SNDLPNLSNKKQKTAASDTKMSFVNSVPSLD--QALDSTKBDQGGTTDNKVRDSEADAK 448
Qy 401 K----DETSSSSSEANSRCOTP-----IKKKPNIPEPENWNSGAEASMRVLIG 445
Db 449 EVGEBEIPNNSVHADGSSICOPHHSGNGAIIIAEMSETSRP--STEMWPIEDLTKGYE 506
Qy 446 TYNDPFCALAR--LIGTCTCRQVYBF--RVKSSIIAPA-----PAEVDV--TP 488
Db 507 IFGRNSCLIANRNLISGLKTLCDVSNMYRNEVSVRBSSTPMLLDDGRTDGNNDNVP 566
Qy 489 PR----KKKKRKLW-----AAHCKRKIQLKKDGSSNNHVNYPQCDHPROPCDSSCPVI 538
Db 567 PTRLFRKRKGRKTKLKYSTKSAGHPSWWRIGAGKQNSCKQYTPCG--CLSMGCKOPCLT 625
Qy 539 AONPEKFCOCSSEBQCNFPFCRC--KAOCNTKQCQCYLAVRECDPDLCLTGCADAHMDK 597
Db 626 NETCCCKCGCSKCNKPRFGCHCAKSCRSQCCPFAAGRCDDPVCNC-----W--- 677
Qy 598 NVSCNCSGI-----QRG-----SKILLAPSDVAGMGIFIKDPVOKNEITSEY 641
Db 678 -VSCGDGSLGAPRPRGEGOCGMRLLAQQQKILLGKSDVAGMGAFILKNSVKNETLGEY 736
Qy 642 CGEIIISODEADRGKYVDKYMCSFLNLNNDVVDATRKANKIRIPANHSVNPNCYAKVM 701
Db 727 TGEIISHHEADRGKITDRANSSFLFDLNDQYVLAQRKGDYKLKPAHNSAKENCYAKWAF 796
Qy 702 VNGDHRIGIFAKRATQTEGELLFPDPRYQADA 733
Db 797 VAGDHRVGI PANERIEASEBELFYDPRYQDDA 828

```

RESULT 3  
T01503  
hypothetical protein T10M13.3 - *Arabidopsis thaliana*  
C1Species: *Arabidopsis thaliana* (mouse-ear cress)  
Accession: F01503  
Date: 24 Mar 1999

CjDate: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #local\_change 24-Mar-1999  
CjAccession: T01503  
R.Johnson, A.F.; de la Bastide, M.; Lodi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; Gott  
Martensen, R.; McCombie, W.  
submitted to the EMBL Data Library, May 1997  
A.Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
Ref: 774346

A:Reference number: Z14348  
A:Accession: T01503  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-898 <J01>  
A:Cross-references: UNIPARC:UPI00000A0AC3; EMBL:AF001308; NID:G2104523; PID:G3912918  
A:Experimental source: cultivar Columbia

A:Map position: 4S  
A;introns: 48/2; 83/1; 118/3; 207/2; 260/2; 288/3; 336/3; 352/3; 554/2; 603/3; 677/2; 7  
A;Note: T10M13.3

Query Match	20.1%	Score 813.5	DB 2	Length 898
Best Local Similarity	28.2%	Pred. No. 1.1e-41		
Matches 248	Conservative 11	Mismatches 270	Indels 221	Gaps 38

[illegible]

QY 242 -----EKKYELTEOOLPGALPBEPTNIDGNPAK-SVOREQ-----SLHSFHTLFCRRC 290  
 DB 283 DASDLIERNELKAD-----KONTEESNSGFKLGISELKGALASFNLCRKL 336  
 QY 291 KYDCFLH-----PFAHTPTTKKKTETALDNKPCPOCYOHLEGAKEPAA-- 337  
 DB 337 VDCRHLGSGOPILSALKAVREVEPC--SNFASKAKEKASEBESKAVSDVPHPAAGA 394  
 QY 338 --LTER-----IKTPKPRG-----GRR-----GRPNNSRP----- 365  
 DB 395 VSLQVEKTDIGIKNVDSGSGVEOEHGIRKREVPILKDSNDLPNLSNKKOKTAASPTKMS 454  
 QY 366 ---SPTTNVLESKOTDSREAGTENGENNKBEKK-----DETSSSEANSRCORP 417  
 DB 455 FNASVPSLD--OALSTKGDGCTTDNKNRDSADAKVGEPIPNDSVHDGSSICOPH 512  
 QY 418 -----IMKPNIBPENVEWSGAEASMFVILGYTDNFCALAE--LITTCRCQ 465  
 DB 513 HSGNGALIIAMSETSRP--STEMNPIEKDLYLKGVEIFGRNSCLIANNLISGLKTCID 570  
 QY 466 VYEF-RVKESSIIADA-----PAEDVD--TPPR-----KKGRHLM-----A 500  
 DB 571 VSNVREKEVSVFRSSSTPNLLLDGRTPDGNNDVEPRTLFRKKGTRKIKYTSKA 630  
 QY 501 AHCRKIOLKGOSSNHNVYVOPCDHPRQPCDSSCPVLAQNFCERKCCOSSSCORPFC 560  
 DB 631 GHPVWKRJAGKNSCKQYTPCG-CLSNCKGDCPLNTECCEKCYGCKSKCKNPRGC 689  
 QY 561 RC-KAQCNTKQCPCTYLAVERCDPLCTTCGADHMDSKNSCKNCSI-----ORG----- 609  
 DB 690 HCAKQCKSRQCPCTYLAVERCDPLCTTCGADHMDSKNSCKNCSI-----ORG----- 609  
 QY 610 -----SKKHLILADSDVAGWGIPIKDPVQKNEPISEYCGEIIISODEADRKVYDKYC 663  
 DB 741 MELLIRLOOORILLGSDVAGWGIPIKDPVQKNEPISEYCGEIIISODEADRKVYDKYC 800  
 QY 664 SFLPLANDPVVDATRKNGKIRFANHSVNPNCYAVVWVNGDHRIGIPAKRAIQGEELF 723  
 DB 801 SFLPLANDPVVDATRKNGKIRFANHSVNPNCYAVVWVNGDHRIGIPAKRAIQGEELF 860  
 QY 724 FDIRYSODA 733  
 DB 861 YDIRYGPDA 870  
 RESULT 4  
 T01127  
 curly leaf protein (polycomb-group) [imported] - Arabidopsis thaliana  
 N/Alternate names: curly leaf protein homolog F26B6.3  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 12-Feb-1999 #sequence\_rev12-Feb-1999 #text\_change 02-Feb-2001  
 C/Accession: T01127; A84624  
 R/Roundsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
 submitted to the EMBL Data Library, June 1998  
 A/Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.  
 A/Reference number: Z14198  
 A/Accession: T01127  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-902 <R0U>  
 A/Cross-references: UNIPARC:UPI00000A76B5; EMBL:AC003040; NID:g3242700; PID:g3242729  
 A/Experimental source: cultivar Columbia  
 R/Lin, X.; Kaul, S.; Roundsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,  
 L.; Neuman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A/Reference number: A84420; MUID:20083487; PMID:10617197  
 A/Accession: A84624  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-902 <STO>  
 A/Cross-references: UNIPARC:UPI00000A76B5; GB:AB002093; NID:g3242729; PIDN:ACC3781.1; G

C/Genetics:  
 A/Gen: At2g23380; F26B6.3  
 A/Map position: 2  
 A/Introns: 21/3; 52/3; 153/2; 207/2; 235/3; 280/3; 296/3; 320/3; 554/2; 604/3; 679/2;  
 Query Match 19.6%; Score 795.5; DB 2; Length 902;  
 Best local similarity 27.0%; Pred. No. 1.4e-40;  
 Matches 251; Conservative 110; Mismatches 274; Indels 295; Gaps 35;  
 QY 9 EKPVCWKRKYSEYRRLRQLKFRRADEVKSMFSSNOKILERTIELIQEWKORRIDEV 68  
 DB 24 ERGPA--SKVSEVIESLKKKLAADRCISIKRIDENKKNLPAIQSFAR----- 71  
 QY 69 HILTVSSLRGRRECVSVDL-----PEPQVITLKLTLN-----AVASY 107  
 DB 72 -----SSMERGSSCKDSDLLVKRQKRSPEKSGSIDSSNNRVYEDGPASSWQGVSSV 125  
 QY 108 PIMYSMSPLQOQFWVEDTVLHNIPLYMGDEVLDQGTPIELILKNYDGKVDHRECFTN 167  
 DB 126 PYSISLRPIYK-----PDIKLSPTTWFILDRQKMTB-----DOSVGR----- 167  
 QY 168 DEIFVELVNALQYVNDDDDDGDDPEERBEKKDLEDRDKESRPPRKF----- 218  
 DB 168 RIYVDQGGBALICSDSEBEBAIDDEEK-----RDLR-----PEDYIIRMTLEQL 213  
 QY 219 -PSDKIFPAISMPDPKGTABRLKAYKELTEOOLPGALPBEPTNIDGNPAKV---OR 274  
 DB 214 GLSDSVLAEILAS-FLSRSTG-BIKRAGVLMKK-----EVSBSGNOAESLNDKM 264  
 QY 275 BOSLSHFHTLFCRRCFKVDCFLH-----PFAHTPTTKKKTETALDNKPCGP 322  
 DB 265 BGALDSFNNLCRKLTVDCRHLGSGODLIPPAEPAWCP-----PVDSULTGA 315  
 QY 333 QCYOHLGAKEPAAALTAE-----RIKTPKPRGRRRGLPNN--SSRSTPTI 370  
 DB 316 NCYKTLKSGRFPQVGTIEGKTGSSDAGKTTPTTKSSKLNGKPKPTFPESASVNEK 375  
 QY 371 NVLES-----KDTSD-----REAGTETGENN----- 393  
 DB 376 CALFTSDENGLQOQTNDSKVSSEPKVYSGRRVGRKKNRVAERVPRTQKQKTEA 435  
 QY 394 -----DKEEBKDKDTSSS-----SEANSRCOTPIKPKP----- 423  
 DB 436 SDSDSIASGSCSPSDAKKDNBDATSSQKIVKSGNSGSKRNGTPAIVSNNSVDDVPV 495  
 QY 424 -----IBPENVE-----WSGAEASMFVILGYTDNFC 452  
 DB 496 QOSNEVASBELAPGSDSLRKEBFMGETVSRGLATNKLMPLEKSLFDKGVIRGNMSC 555  
 QY 453 AIAE--LIGTYTCQVYEFV-----KESIIAPAPADVDTTPPKKGRK 495  
 DB 556 LIAEHLISGFKSCVEFQYMTCESEKASFPFGDGLNPDGSKFTDINGMNVNQVARRSRF 615  
 QY 496 HR-----LW-----AAHC--RKIQKXGSSNHNVYVOPCDHPR-----PDS 532  
 DB 616 LRRGKVRRLTYTWSAAYHSIRKITEKKO-----QPC--KQFNPCKCIAQK 663  
 QY 533 SCPCVIAQNPCEKCCOCSSECONRPPGRC--KAQCNTKQCPCTYLAVERCDPLCTTCGA 591  
 DB 664 BCPGLNLTGCEKTCGCKSKCNFRGCHCAKSGQRSGQCPAFADRCDDPVCNC----- 720  
 QY 592 DHW-----DSKNVSCNKSIOGSKKHLILAPSDVAGWGIPIKDPVQKNEFI 638  
 DB 721 --WYIGDGLGVDSORDDNVECNMKLLKQOQRVLLIGISDVGWGFLLKNSVSKHYL 778  
 QY 639 SEYGEIISODEADRGVYVQKWCSPFLPNLNDPVVDATRKNGKIRFANHSVNPNCYAK 698  
 DB 779 GEYTGELISHREADRGKTYDRENCSPFLPNLNDPVVDATRKNGKIRFANHSVNPNCYAK 838  
 QY 699 VMMVNGDHRIGIPAKRAIQGEELFFDYRY 728  
 DB 839 VIMVAGDHRVIGIPAKERLAGEELFYDRIY 868

## RESULT 5

T52060  
Protein MEDA [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004  
C/Accession: T52060  
R/Grossniklau, U.; Velle-calzada, J.P.; Hoepfner, M.A.; Gagliano, W.B.  
Science 280, 446-450, 1998  
A/Title: Maternal control of embryogenesis by MEDA, a polycomb group gene in Arabidopsis  
A/Reference number: Z55927; MUID:96212068; PMID:9545225  
A/Accession: T52060  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-689 <GRO>  
A/Cross-references: UNIPROT:O65312; UNIPARC:UPI000003C2EE; EMBL:AF060485; PIDN:AAC39446.  
C/Genetic8:  
A/Gene: MEA

Query Match 17.9%; Score 726; DB 2; Length 689;  
Best Local Similarity 27.3%; Pred. No. 1.7e-36;  
Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

```
QY 120 FMVEDTVLHNI PYMGDEVLDODGTFT---BELIKNYDGKYGHDRCGFINDILFYELV 175
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 114 FLDEVDPLPSVKLPYVKLPRLSTITWFTKSSQLMAESD-SVIGKQIYINGE----- 166
QY 176 NALQGYNDODDDDDGDDDEE-REBKQKDLHDHDDKSRPPRKPSDKIFEAISMFPDK 234
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 167 -ALBISSEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEBDEB 208
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 235 G-----TAEIKEXKXELTEQOLPGALPPBCTPNIDPNKASVQREQL 278
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 209 GLDULVVARALAKYLEVNVSDILBRYNELKLN-----DGTAGAA--SDLTS 253
QY 279 HSFHTLF-----CRRCFKYDCFLPFPATPTNTYKKNKTETAL---DNKPCGQCTQH 327
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 254 KTIITAFQDFADRRHRCRCMIJFDCMHKRY-EPERSSEBDSKSLFEDEBRQPCSEHCYIK 312
QY 328 LEGAKFPAALTAERIKTPPKRPGRRGRRLPNSSRPSTPTINLBSKQDTSDEAGTE 387
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 313 VRSVTE-----ADNV-----MDNDS----- 328
QY 388 TGGENNDEBEEKDEKDETSSEANSRCQTPIMKKEINIBPENWMSGAEAMFRVLIGTY 447
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 329 -----ISNKIVGDDPNNTMTTPYBKOLYLKGIELF 358
QY 448 YDNFCAIAR-LIGTTCRQYEFVYKESST-----IAPAPAEVD 486
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 359 GRNSCDVALNLRGLKTCLEIYNVREDOQCTMSLDLNTQORHNOVTKVSRKSSRSV- 417
QY 487 TPRPKKKKRLMAHCKRIQKQSSNHYVNYPCDHPROPSCSPCYAONFCSEKF 546
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 418 ---RKSKRLKRYAYPPALKTTTSGEAKFYKHYTPCT-CKSKCGQCCPLTHNECKEY 472
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 547 CQCSBECQNRFPQCRCK-AQCNTKQCPCTYLAVRECDPDLG-----LTCGAADHMSK-NVS 600
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 473 CQCSGDCQNRFPQCGCAIGQCTNRCQPCFPAANRECDPDLGSCGGLTGERPVQIQ 532
QY 601 CKNGSIQSGSKKHLILAPSDVAGMGIPTKDPVQKNKFISEYCGEIIISQDEADRRKGVYDK 660
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 533 CKNGQFLLQTNKKIILIGKSDVHGMGAFTWDSLKKNKEYLGEYTGELITDEANERERIDR 592
QY 661 YMCSTLFLNANDPVVDATRKGNKIRIPANSHVNPNCYAKVMVNGDHRIGIPAKRAIOTGE 720
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 593 IGSSTLFTLNDQLEIDARKGNBFFGLNHSANPNCYAKMLTVRGDRIGLFPAERALBGE 652
QY 721 ELFFDYRYQADA 733
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 653 ELFFDYCYGRBHA 665
```

RESULT 6

T21436  
hypothetical protein R06A4.7 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C/Accession: T21436, T23953  
R/Barlow, K.  
Submitted to the EMBL Data Library, November 1996

A/Reference number: Z19421  
A/Accession: T21436  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-775 <WLS>  
A/Cross-references: UNIPARC:UPI000017BAPD; EMBL:Z81515; PIDN:CAB04199.1; GSPDB:GN00020;  
A/Experimental source: Clone F26H11  
R/Barlow, K.  
Submitted to the EMBL Data Library, November 1996

A/Reference number: Z19822  
A/Accession: T23953  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-775 <W12>  
A/Cross-references: UNIPARC:UPI000017BAPD; EMBL:Z83120; PIDN:CAB05589.1; GSPDB:GN00020;  
A/Experimental source: Clone R06A4  
C/Genetic8:  
A/Gene: CESP:R06A4.7  
A/Map position: 2  
A/Introns: 223/1; 414/3; 464/2; 537/1; 601/2; 655/1; 679/1; 749/2

Query Match 15.3%; Score 618; DB 2; Length 775;  
Best Local Similarity 26.4%; Pred. No. 6.8e-30;  
Matches 180; Conservative 89; Mismatches 244; Indels 170; Gaps 25;

```
QY 106 SVPIWMSPLQGNFMVEDTVLHNI PYMGDEVLDODGTFTBELIKNYDGKYNH-DRECG 164
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 170 SFPWLY-WPIEGVATPDDQLRLTHMPYFGDGI--DDGNLYEHLIDWFPPIGHSFNWS 226
QY 165 FINEIFVELNALGQVNDODDDDDGDDPERREBKQKDLHDHDDKSRPPRKPSDKIF 224
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 227 YVNDWILYKLCRAA-----LKDYQGS-----PDVY 252
QY 225 EAISSMFPDKGTAEIKEXKXELTEQOL-----PGALPP-----ECTPNIDPNKASV 272
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 253 YTLRYLWPNKSSQGFSSAPFVLCENFAKGFDDSSLEPMKKTIAGAGQVLRAPT----- 308
QY 273 QREQLSHSFHTLFCRRCFKYDCFLPFPATPTNTYKKNKTETAL---DNKPCGQCTQH 315
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 309 -----CYACIATYCAIHGFABIRIPNGEFYNAMLPLPN-----NPEV-- 348
QY 316 DNKPCGQCTQHLEGAKFPAALTAE-----RIKTPPKRPGRRGRRLP 359
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 349 DGKMSGNCKWKSAT-MKEVSEVLVPSDEBELQKEVKIYFMKSRLAKMFIEDGALIVNIV 407
QY 360 NNS---SRPSTPTINLBSKQDTSDEAGTETGGENNDEBEEKDEKDETSSEANSRCQTP 417
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 408 FNTIIVAPRCFV-----KKYVDEBDEBSKIRSCDAYHLM 443
QY 418 IKMKRNTI-----PRENVMSGAEAMFRVLIGTYDNFCAIARLIGTTCRQYEF 470
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 444 MSMAENVASRLKMGQPSN-----RLSFKDRVNNRR-----NQLSQEKAKRK 486
QY 471 VKESIIAPAPAEVDTPPRKKKKRRLMAHCKRIQKQSSNHYVNYPCDHPROPSC 530
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 487 LRHDSLRLQALRDGLDEKLIREDDMRSSQNSSEKVRN-----TAVVPTACRH-AGPC 539
QY 531 DSSCP-CYIAON-FCEKPCQCSBECQNRFPQCRCKA-QCNTKQCPCTYLAVRECDPDLG 587
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 540 NATANCKACRENGCSTYCKCDINCSQRFPCGNCALAGCTYKACQCFPAANRECDP 599
QY 588 CGAADHMSDKVNSCKNGSIQSGSKKHLILAPSDVAGMGIPTKDPVQKNKFISEYCGEIIIS 647
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 600 C-KCDALDSNIIKCNFNFGMTBIMIQRTYCGPSKIAGNGLFLLEPAEKEDEFTYTGERTIS 658
QY 648 QDEADRRKGVYDKTMCSFLFLNANDPVVDATRKGNKIRIPANH-SVNPNCYAKVMVNGDH 706
```



DB 659 DDEAERGAIDYRQCYIFNIETGALIDSYKIGNLARFANDHSKNPTCTARMTVAGSH 718  
 QY 707 RIGIFAKRAIQCEBELFPDYRS 729  
 DB 719 RIGFYAKRRLRISBELTFDYSYS 741

## RESULT 7

S71490  
 ashl protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C>Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 09-Jul-2004  
 C:Accession: S71490; S78442  
 R:Triopoulos, N.; Ledemese, D.; Gilda, J.; Shearn, A.  
 Genetics 143, 913-928, 1996  
 A>Title: The Drosophila ashl gene product, which is localized at specific sites on polyt  
 A:Reference number: S71490; MUID:96363924; PMID:8725238  
 A:Accession: S71490  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2144 <TR1>  
 A:Cross-references: UNIPROT:Q24189; UNIPROT:Q9VWA15; UNIPARC:UPI000017BE7A; EMBL:U49439  
 R:Triopoulos, N.; Shearn, A.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: S78442  
 A:Accession: S78442  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-119, 'V', 121-611, 'X', 613-2144 <TR2>  
 A:Cross-references: UNIPARC:UPI000017BE7B; EMBL:U49439  
 C:Genetics:  
 A:Gene: ashl  
 A:Introns: 1592/3, 2083/1

Query Match 7.6%; Score 306.5; DB 2; Length 2144;  
 Best Local Similarity 22.9%; Pred. No. 1.6e-10;  
 Matches 141; Conservative 70; Mismatches 200; Indels 205; Gaps 26;  
 QY 241 KEKYELTEQULPAL---PPECTPNIDGNPAKSVQREGSLHSFHT---LFCRCF 290  
 DB 889 RKRKRVSEQQTAAVVDHEHPFDP--DDEPLQSLRETSNNVAVQAAPPLDCER-- 944  
 QY 291 KYDCLHPHATPPNYKKNKTETALDNKCCGQCQHLGAGAPAAALTAETIKTPPKRP 350  
 DB 945 -----VPOGAEATFVARTNOK-APR-----LSVALERLQRPQTPA 981  
 QY 351 GGRRGRLPNNSR-----PSTPTINVL-----SEKOTDSREAGTERGENNDKE 396  
 DB 982 RGRPRGRKPKRQQAARQPPKSEPEIRPAKKGKGRQKQVLEBPPTPPQKKNM 1041  
 QY 397 EEEKDETSSESSSEANSRCQTPIKMKENIE-----PPE---NVMSG 434  
 DB 1042 EPNILPQIDPNTFSCIRLRKRNLEAGTQPKKQVQVTVBEIPPEIPVQGEID 1101  
 QY 435 AEASFRV-LIGTYDNRFA-----IARLIGTKCR 464  
 DB 1102 AEAEKRVDSIPTEDPLPASSESHPGPDVASCSESEDKASTTSLRKLKVKKTYLVA 1161  
 QY 465 QVVEPRVSESLIAPA-----PAEDVTPPKKKKKR-----LWAA 501  
 DB 1162 GLSHHYQSLMPAPKNNKKGLBEQVGPASLPPPEYCEKYLARTEMDFLPDIWMA 1221  
 QY 502 HC-----RKIQKQSGSSNHY-----NYOPCDHPROPDSSCPVTAQ 540  
 DB 1222 YNSKLPTRNVPVSNWYRKIR-----TNVAASVRPLAGFDPH-----TNC- -K 1265  
 QY 541 NFCEKCCGSSRCQRFRCRCKAQNTKQCCYLAVECPDCLCTGGAADHMSKVS 600  
 DB 1266 NOGK--SCLDNCNR-----NVYTESPS--NCPAGE-----K 1295  
 QY 601 CKNCISIQSGSKHLLAP-----SDVAGWGFIDPQVQKNFISYCGEIIISODE-ADR 653

DB 1296 CENQIKORHA-----VAPGVERFMTADKMGVTRKLP.IAKGTYLEYVEGVTEKEFKOR 1350  
 QY 654 RGVYDKTMCSEFLNANDFVVDAIRKKNIRFPAHNSVNPNCYAVMANGHRIGIPAK 713  
 DB 1351 NASIYNDTHNYCHLDDGLVLDGQRMGSDCFVNHSCBPNCEQMKWSVNGISRYVLPFAK 1410  
 QY 714 RAIQGEELFPDYRS 729  
 DB 1411 RAIBGEELTYDYNFS 1426

## RESULT 8

A44265  
 trithorax homolog HTX, version 2 - human  
 M:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: A44265; A44264; I58112; I37165; I38485  
 R:TKachuk, D.C.; Kohler, S.; Cleary, M.L.  
 Cell 71, 691-700, 1992  
 A>Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translo  
 A:Reference number: A44265; MUID:93046667; PMID:1423624  
 A:Accession: A44265  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-3968 <TRK>  
 A:Cross-references: UNIPROT:O03164; UNIPARC:UPI000017C3FD; GB:L04284; NID:G184393; PID  
 A>Note: sequence extracted from NCBI Backbone (NCBI:117729)  
 R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canaan, O.; Cimino, G.; Croce, C.M.; C  
 Cell 71, 701-708, 1992  
 A>Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1  
 A:Reference number: A44264; MUID:93046668; PMID:1423625  
 A:Accession: A44264  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 63-316, 'GLINSELEK', 327, 'Q', 329, 'VR', 332, 'DKEGTP', 340, 'T', 342, 'EDKTVVRQSP;  
 546, 'LQIRSTSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GUI>  
 A:Cross-references: UNIPARC:UPI000017C3FE; GB:L04731; NID:9339921  
 A>Note: sequence extracted from NCBI Backbone (NCBI:117779)  
 R:DiDall, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.  
 Nature Genet. 2, 113-118, 1992  
 A>Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in ac  
 A:Reference number: I58112; MUID:93265134; PMID:1303259  
 A:Accession: I58112  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328 <DNA>  
 A:Cross-references: UNIPARC:UPI000016B134; GB:L01966; NID:G307522; PIDN:AAA92511; PI  
 R:Matuchalet, R.; Grell, J.; Lochner, K.; Nilsson, I.; Stiegler, G.; Zwechbrunner, I.; B  
 B. J. Haematol. 90, 308-320, 1995  
 A>Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in th  
 A:Reference number: I37165; MUID:95315013; PMID:7794749  
 A:Accession: I37165  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1212-1603, 'GTR', 'MAR>  
 A:Cross-references: UNIPARC:UPI000016A531; EMBL:X83604; NID:G897757; PIDN:CA58584.1; I  
 A>Note: submitted to the EMBL/GenBank/DBJ databases by R. Matuchalet, 20 December 199  
 R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canaan, O.; Saito, H.  
 Cancer Res. 54, 2327-2330, 1994  
 A>Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved  
 A:Reference number: I38485  
 A:Accession: I38485  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1251-1486, 'G', 1488-1538 <RS>  
 A:Cross-references: UNIPARC:UPI000016A07B; EMBL:U04737; NID:G451554; PID:G451555  
 C:Genetics:  
 A:Gene: GDB:MLI, HTX, ALL-1, HRX  
 A:Cross-references: GDB:128819; OMIM:159555  
 A:Map position: 11q23-11q23  
 A:Introns: 1338/1, 1362/3, 1406/3, 1444/3, 1493/3, 1525/3, 1566/1  
 A>Note: the list of introns is incomplete

C:Superfamily: histone methyltransferase, trithorax protein type  
 C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger  
 F:1434-1456/Region: zinc finger CCHC motif  
 F:1479-1506/Region: zinc finger CCHC motif  
 F:1527-1556/Region: zinc finger CCHC motif  
 F:1569-1596/Region: zinc finger CCHC motif  
 F:1673-1900/Region: zinc finger CCHC motif  
 F:1933-1955/Region: zinc finger CCHC motif

Query Match 7.4%; Score 300; DB 2; Length 3968;  
 Best Local Similarity 19.7%; Pred. No. 7.7e-10;  
 Matches 131; Conservative 86; Mismatches 203; Indels 246; Gaps 23;

207 RDDKSRPPKRPSPDKIFEAISMFPPDKGTAE----- 238  
 3392 QDQVVALPPS-----SGMFPQLGTSQTPSTAATAASSICVLPSTQTGTAAAS 3440  
 239 ---ELKEKYK-----ELTEQQLPGALPPPC---TNIDGPNAKSVQREBSL 278  
 3441 PSGEADERTYQLQHVNOVLASTKTHISQORDLSASGPQVSNFTQTVAPNSMGLQNKAL 3500  
 279 HSFHTLFCRCFCFKYDCFLHPHATPNTYKRNKTETALDNKPCGQCQYHLEGAKFPAAAL 338  
 3501 SS-----AVQASP-----TSFGSSPSFSQGR-----S 3524  
 339 TAERIKTPPKRPGRRGLRPNNSRPSTPTINYLESKOTDS---DREAGTGTGENDK 395  
 3325 ASPVPQPTKPKPKTKRQFLDLKNGKKGKHYSHLRTSSSEAHIPDQTTSLTSGTTPG 3584  
 396 EEEKQDETSSSSSEANSRCQTP---TKKKNIE-----PPNNV----- 431  
 3585 AEACQODTASILEOSQKECGPAGQVAVLPEVQYTONPANEQESAPKTVESBSNFSPP 3644  
 432 ---WGAAS-----MFRV----- 442  
 3645 LMLWLOEQEQRKKESTTEKKPKKGLVFEISSDDGFOICAESIEDAMKSLTDKVOEARSNAK 3704  
 443 -----LIGTYDNYFC-AIARLIGTKTCROYVEFRVYKSSIIAPAPAEVDVTP 488  
 3705 LKQLSFVAGNGLRMLGIHLDAVFLIEQLSGAKHCN-YKFRPHK-----PEANEP 3755  
 489 PRKKRKRRLMAACRKLQLKDGSSNNHYNYOPCDHPROPDSSCPVIAQNKCEKQ 548  
 3756 PLNPHGSARA-EVHLRK-----SAFMFNLASKH-RQP-----PEYNPNDEEBEVQ 3801  
 549 CSSECONRPFPGCRCAQCNTRKQPCYLAVRECDPDLCTCGAADHWDKSNVSCNKSICIOR 608  
 3802 LKS-----ARRATSMDLPMRPFH-----LKK 3824  
 609 GSKHLLAPSDVAGWGIPIFDPOVQKNEFISEYCGEIIISODEADRGRKYVD-KYMCSEFLP 667  
 3825 TSKEAVGVYRPIHGRGLFCRKNIDAGEMVIEVAGNVIRSIQTDKREKCYYSKIGCYMF 3884  
 668 NLANDFVVDATRKNGKIRFANHSVNPNCYAKVMVNGDRIRGIFAKRAIQTGELFPDYR 727  
 3885 RIDSEVVDATMGHRRARFINHSCBNCTSRVINIDQKHIVIFAMRKTYRGEELTYDYK 3944  
 728 YSQADA 733  
 3945 PPIEDA 3950

# RESULT 9

A48205  
 ALL-1 protein +GTR form - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 22-Jun-2003  
 C:Accession: A48205; B48205  
 R:Ma, Q.; Alder, H.; Nelson, K.K.; Chatterjee, D.; Gu, Y.; Nakamura, T.; Canaan, E.; Ch  
 Proc. Natl. Acad. Sci. U.S.A. 90, 6350-6354, 1993  
 A:Title: Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 an  
 A:Reference number: A48205; MUID:93317679; PMID:8327517  
 A:Accession: A48205

A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-3869 <HAR>  
 A:Cross-references: UNIPARC:UPI000017C621; GB:L17069  
 A:Accession: B48205  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1505,1509-3869 <HA2>  
 A:Cross-references: UNIPARC:UPI000017C622; GB:L17069  
 C:Genetics:  
 A:Gene: All-1  
 C:Superfamily: histone methyltransferase, trithorax protein type  
 C:Keywords: alternative splicing; zinc finger

Query Match 7.4%; Score 298; DB 2; Length 3869;  
 Best Local Similarity 21.2%; Pred. No. 9.8e-10;  
 Matches 139; Conservative 84; Mismatches 206; Indels 228; Gaps 25;

207 RDDKSRPPKRPSPDKIFEAISMFPPDKGTAE----- 238  
 3393 QDQVVALPPS-----SGMFPQLGTSQTPSTAATAASSICVLPSSQTAGMTAAAS 3341  
 239 ---ELKEKYK-----ELTEQQLPGALPPPC---TNIDGPNAKSVQREBSL 276  
 3342 PSGEADERTYQLQHVNOVLASTKTHISQORDLSASGPQVSNFTQTVAPNSMGLQNKAL 3401  
 277 HSFHTLFCRCFCFKYDCFLHPHATPNTYKRNKTETALDNKPCGQCQYHLEGAKFPAA 336  
 3402 TLPSAKP-----ASAPSGSSPSGQSGSSSVF-GPT-----KP 3435  
 337 ALTKERIKTPPKRPGRR-----RGLRPNNSRPS-----TPTINYLESKOT- 378  
 3436 KPKAKRIQLPDKGSYKHKVSHLRTSSSEAHIPRDPPAPQPSVTTPPAN-REQDAA 3494  
 379 ---DSREAGTENG-----GENNDKEBEKK---DETSSSEANSRCQTP 418  
 3495 GVEQPSQKEGCGPAPVPAALPEVQATQNPANEQENAPKAMEEBSGFSPLMLWLOEQ 3554  
 3555 KRKESITERPKPKGLVFEISSDQFOICAESIEDAMKSLTDKVOEARSNAKLSFVAGV 3614  
 419 KMKENI---BPENVEWSGAASMFV----- 442  
 443 -----LIGTYDNYFC-AIARLIGTKTCROYVEFRVYKSSIIAPAPAEVDVTPPKRKRKH 497  
 3615 NGLMGLIHLDAVFLIEQLAGACHCN-YKFRPHK-----PEANEPPLNPHGSAR 3665  
 498 LMAACRKLQLKDGSSNNHYNYOPCDHPROPDSSCPVIAQNFCEKFCQCSSECONRF 557  
 3666 A-EVHLRK-----SAFMFNLASKH-RQP-----PEYNPNDEEBEVQQLS----- 3705  
 558 PGCRCAQCNTRKQPCYLAVRECDPDLCTCGAADHWDKSNVSCNKSICIORGSKHLLA 617  
 3706 ---ARRATSMDLPMRPFH-----LKKTSKEAVGVY 3724  
 618 PSDVAGWGIPIFDPOVQKNEFISEYCGEIIISODEADRGRKYVD-KYMCSEFLNANDVVD 676  
 3735 RSPHIGRGLFCRKNIDAGEMVIEVAGNVIRSIQTDKREKCYYSKIGCYMFRIDSEVVD 3794  
 677 ATRKNGKIRFANHSVNPNCYAKVMVNGDRIRGIFAKRAIQTGELFPDYRYSQADA 733  
 3795 ATMGNAARFINHSCBNCTSRVINIDQKHIVIFAMRKTYRGEELTYDYKPIEDA 3851

# RESULT 10

T41282  
 probable transcription silencing protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T41282  
 R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, April 1999  
 A:Reference number: T21984  
 A:Accession: T41282





Query Match 6.9%; Score 279; DB 2; Length 1767;  
 Best Local Similarity 20.0%; Pred. No. 5.8e-09;  
 Matches 172; Conservative 118; Mismatches 283; Indels 286; Gaps 37;

QY 1 MGQTGKSEKGFVCKRKVSSYMLRLQKRRPAD-EVYKSPSSNRQKILERTILNCE 59  
 DB 441 MSSSRKADGQCKVQTKTSARHLRKSSRKQSRDIIESIFKCSKQK-RSSILK-- 495  
 QY 60 WKQRIQPHILTVSLSLACTRECVSLDLPPTQVPIPKTLNAAVASIVIMSMELQON 119  
 DB 496 -----TSRSSEWG-----LPKTYEIR-----LQSN 516  
 QY 120 FMVEDEVLANIPYNG---DEVLDQGTPI-BELIKYDGVKVGDR-----EC--- 163  
 DB 517 -----NIPDGPPIHEPORSOGNLANGEHRSSNANGVSGSNRIQASSGCLRL 566  
 QY 164 -----GFINDELFVELVNALGOYND-----DDD 187  
 DB 567 KYKPGKGGQNPNTIVTSKVGNSLPGNGIVKAGTCLBLPGS-AHGEDMKQTVETKEDL 625  
 QY 188 DDGDDPEREREKQKOLEHRDOKESRPP---RKPPSDKIFP-AISSFPDQKTAELK 241  
 DB 626 VEKSPVEKVSYLQSSDSNRDKKYNQDAGLCRKVGVDLDDPHLSI-----RMV 677  
 QY 242 EYKKEUTEQQLPGALPPECTPNIDGPN-----KSVQREOSLHSPHTLFCRRCFKYDCL 296  
 DB 678 BECEBATGTQ---SLDAETSPSEVINSPDSIVITIEHSEGLH----- 717  
 QY 297 HEPHATPNTYKRNKTETALDNKPCGPOCYOHLGAKAFALALT-----AERIKT 345  
 DB 728 HGFSTPDEVVKN-----RVLEKDELPASKSPSNGSHLIPNKKAKH 762  
 QY 346 PPKRPGRRRG-----LPNNSRPST--PTIVVLESKDT- 378  
 DB 763 PPSKNGTKKSKSESADGRKNSHEGVQRKSLNTSMGRDSDYPEVGRISHKTT 822  
 QY 379 ----DSDBAGTETGGE--NNDKEBEKK-----DETSSSBSANSRQPIIKKPIETPE 428  
 DB 823 GALLDADIGKTSATITGSSDVTGHBMYVDVTEISYSTESAMVRCDDCFMR---RIPA 879  
 QY 429 NVEWGAESAEMFRVLIGTYDNFC---AIALIGTKTCQVYEFVKESSIIA--PAPAE 483  
 DB 880 SVVGSIDESSRW-----ICMNSDKRFADCSKQSEMNERINEBELIGQDEADAY 929  
 QY 484 DVTTPPKKKKRRHMAAHCKRIQLKQSSNHVNYTOPCHPR--QPCDSSCPVIAON 541  
 DB 930 DCDARKGKEKQK-----SKRLTGKQKACFVAKITNOFLHRRKRSQTIIDELMVC----- 979  
 QY 542 FCEKFCQ-----CSSECONRPPGCRCKAQCNTKQCPYLAVRECPDLCLTGADH 593  
 DB 980 ----HCKSPDGRLOGCEBCLNMLNTEC-----LOGTCPAGD- 1013  
 QY 594 WDSKNVSCNCSIQRSKXKHLALAPSDVAGWGIPIKDPYQKNEPLSEYCGEIIISQDEAD 653  
 DB 1014 -----LCSNQOQKARKYKFERFOSGKGYGLRLLEVRBGOFLIEYGVGLDMQSYET 1067  
 QY 654 RGVTV---DKTMCSPFLFNLANDFYVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIG 709  
 DB 1068 RQKEVAFKQKIF--YFMTLNGNEVYIDAGAKNIGRFINHSCEPNCRTKMMVNGEICVG 1125  
 QY 710 IFAKRAIQGBELFPDYRY 728  
 DB 1126 IFSMODLKKGDELTDPYNT 1144

## RESULT 15

T47820  
 Hypochemical protein P24G16.230 - Arabidopsis thaliana  
 C1Species: Arabidopsis thaliana (mouse-ear cress)  
 C1Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
 C1Accession: T47820  
 R1D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;  
 submitted to the Protein Sequence Database, February 2000

A:Reference number: 224477  
 A:Accession: T47820  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-352 <DNA>  
 A:Cross-references: UNIPROT:Q9M1X9; UNIPARC:UPI000009D135; EMBL:AL138647  
 A:Experimental source: cultivar Columbia; BAC clone P24G16  
 C:Genetics:  
 A:Map position: 3  
 A:introns: 6/3; 47/1; 84/2; 117/3; 144/1; 201/2; 228/2; 278/3; 312/3; 329/3  
 A>Note: P24G16.230

Query Match 6.8%; Score 275.5; DB 2; Length 352;  
 Best Local Similarity 29.6%; Pred. No. 1.6e-09;  
 Matches 72; Conservative 32; Mismatches 78; Indels 61; Gaps 6;

QY 505 KIQLKQDSSNHVNYOPCDHROP-----CDSSCPVIAONFCBEKFCQSSSECONRPPG 559  
 DB 51 KKKLKKVKYKHGIF-----CSCSLDPGSSTLGSDCNCGILLSSCSSCKSCSECTNK--- 103  
 QY 560 CRCAQCNTKQCPCYLAVRBCDPDLCLTGADHMDSKVSCNCSIQRSKXKHLALPS 619  
 DB 104 -----PQQRHKKKKLYOT 118  
 QY 620 DVAGWGIPIKDPYQKNEPLSEYCGEIIISQDEADR--GKYVDKYMCSF--LFNLNDFYVDA 677  
 DB 119 EKCQGIYADEIDNSGEPFLIEYGVLDKICEERLMTLNHKEVTNFTLCQINMMWVIDA 178  
 QY 678 TRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQGBELFPDYRYSQ--ADALK 735  
 DB 179 THKGNKSYINHSCSPNTEWQKWIIDBETRIGIFATRIINNGEOLTYDYQFQFADQDC 238  
 QY 736 YVG 738  
 DB 239 YCG 241

Search completed: August 11, 2006, 22:29:06  
 UDB time : 38.427 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:17:04 ; Search time 199.794 Seconds

(without alignments)  
3453.870 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051

Sequence: 1 MGQCKSEKBPVCMKRVK.....RYSQADAKVYGIEMKRLP 746

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Databse : UniProt 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4051	100.0	746	1 EHZ2_HUMAN	Q15910 homo sapien
2	4046	99.9	746	2 Q4R381_MACFA	Q4R381 macaca fasc
3	4030.5	99.5	751	2 Q96F16_HUMAN	Q96F16 homo sapien
4	3999	98.7	746	2 Q99L74_MOUSE	Q99L74 mus musculu
5	3999	98.7	779	2 Q571L5_MOUSE	Q571L5 mus musculu
6	3993	98.6	746	1 EHZ2_MOUSE	Q12188 mus musculu
7	3990	98.5	746	2 Q3YZH6_MOUSE	Q3YZH6 mus musculu
8	3922	96.8	742	2 Q6XKH7_MOUSE	Q6XKH7 mus musculu
9	3809	94.0	748	2 Q5XKH0_XENTLA	Q5XKH0 xenopus lae
10	3793	93.6	748	2 Q96SM3_XENTLA	Q96SM3 xenopus lae
11	3758	92.8	748	2 Q4V8M3_XENTLA	Q4V8M3 xenopus lae
12	3698.5	91.3	695	2 Q4R780_MACFA	Q4R780 macaca fasc
13	3631	89.6	664	2 Q75MP9_HUMAN	Q75MP9 homo sapien
14	3610.5	89.1	669	2 Q75M00_HUMAN	Q75M00 homo sapien
15	3394.5	83.8	759	2 Q90WPM_TETNG	Q90WPM tetraodon n
16	3369	83.2	760	2 Q5TKR5_ORYLA	Q5TKR5 oryzias lat
17	2606.5	64.3	747	1 EHZ1_MOUSE	P70351 mus musculu
18	2606.5	64.3	750	2 Q3U3V5_MOUSE	Q3U3V5 mus musculu
19	2606.5	64.3	750	2 Q922L1_MOUSE	Q922L1 mus musculu
20	2603.5	64.3	747	1 EHZ1_HUMAN	Q92800 homo sapien
21	2603.5	64.3	747	2 Q53XP3_HUMAN	Q53XP3 homo sapien
22	2592.5	64.0	747	2 Q5RDS6_PONPY	Q5RDS6 pongo pygma
23	2421.5	59.8	782	2 Q4S976_TETNG	Q4S976 tetraodon n
24	2278.5	56.2	683	2 Q3UW02_MOUSE	Q3UW02 mus musculu
25	2156.5	53.2	760	1 BZ_DROME	P41124 drosophila
26	2148.5	53.0	742	2 Q7PT13_ANOGA	Q7PT13 anopheles g
27	2105	52.0	749	2 Q2LZJ3_DROPS	Q2LZJ3 drosophila
28	2082	51.4	508	2 Q504B3_BRARB	Q504B3 brachydanio
29	2048	50.6	391	2 Q3U575_MOUSE	Q3U575 mus musculu
30	1768	43.6	376	2 Q59H64_HUMAN	Q59H64 homo sapien
31	1631.5	40.3	579	2 Q8BR85_MOUSE	Q8BR85 mus musculu

32	1631.5	40.3	582	2 Q3TPR1_MOUSE	Q3TPR1 mus musculu
33	1515	37.4	286	2 Q6R125_HUMAN	Q6R125 homo sapien
34	1290.5	31.9	354	2 Q4T531_TETNG	Q4T531 tetraodon n
35	1261.5	31.1	346	2 Q4T5B4_TETNG	Q4T5B4 tetraodon n
36	1163.5	28.7	231	2 Q8C215_MOUSE	Q8C215 mus musculu
37	1141.5	28.2	312	2 Q4SVV6_TETNG	Q4SVV6 tetraodon n
38	1136	28.0	395	2 Q4TFJ6_TETNG	Q4TFJ6 tetraodon n
39	1105	27.3	247	2 Q6GMT4_XENTLA	Q6GMT4 xenopus lae
40	1080	26.7	269	2 Q4TUB3_TETNG	Q4TUB3 tetraodon n
41	1026.5	25.3	267	2 Q4TF58_TETNG	Q4TF58 tetraodon n
42	918	22.7	246	2 Q4TPK7_TETNG	Q4TPK7 tetraodon n
43	842	20.8	854	2 Q76194_PETHY	Q76194 petunia hyb
44	828.5	20.5	816	1 EZA1_ARATH	Q92588 arabidopsis
45	828.5	20.5	916	2 Q76195_PETHY	Q76195 petunia hyb

#### ALIGNMENTS

RESULT 1  
ID EHZ2\_HUMAN STANDARD; PRT; 746 AA.  
AC Q15910; Q15755; Q92857;  
DT 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.  
DT 15-JUL-1998, sequence version 2.  
DT 07-FEB-2006, entry version 44.  
DE Enhancer of zeste homolog 2 (ENX-1).  
GN Name=EHZ2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
OC Homo.  
NCBI\_TaxID=9606;  
RX MEDLINE=97124843; PubMed=8954776; DOI=10.1006/geno.1996.0588;  
RC Tissue=Brain;  
RP NUCLEOTIDE SEQUENCE.  
RA Chen H., Rosster C., Antonarakis S.E.;  
RT "Cloning of a human homolog of the Drosophila enhancer of zeste gene (EHZ2) that maps to chromosome 21q22.2.";  
RL Genomics 38:30-37 (1996).  
[2]  
NUCLEOTIDE SEQUENCE.  
RX MEDLINE=97357309; PubMed=9214638; DOI=10.1093/emboj/16.11.3219;  
RA Liable G., Wolf A., Doth R., Reuter G., Nislow C., Lebererger A.,  
RA Popkin D., Pillus L., Jenuwein T.;  
RT "Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres.";  
RL EMBO J. 16:3219-3232 (1997).  
[3]  
NUCLEOTIDE SEQUENCE OF 134-746.  
RX MEDLINE=96220494; PubMed=8649418;  
RA Hobert O., Jallat B., Ullrich A.;  
RT "Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression.";  
RL Mol. Cell. Biol. 16:3066-3073 (1996).  
[4]  
ATRX BINDING.  
RX MEDLINE=98167853; PubMed=9499421; DOI=10.1093/hmg/7.4.679;  
RA Cardoso C., Timbit S., Villard L., Khrestchatsky M., Fontes M.,  
RA Colleaux U.;  
RT "Specific interaction between the XNP/ATRX-X gene product and the SET domain of the human EHZ2 protein.";  
RL Hum. Mol. Genet. 7:679-684 (1998).  
[5]  
PHOSPHORYLATION SITE THR-487.  
RX PubMed=15302935; DOI=10.1073/pnas.0404720101;  
RA Beausoleil S.A., Jedrychowski M., Schwartz D., Elias J.E., Villen J.,  
RA Li J., Cohn M.A., Cantley L.C., Gygi S.P.;  
RT "Large-scale characterization of HeLa cell nuclear phosphoproteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:12130-12135 (2004).  
!- FUNCTION: May be involved in the regulation of gene transcription

CC	- chromatin structure.
CC	- SUBUNIT: Binds ATRX via the SET domain (Probable).
CC	- SUBCELLULAR LOCATION: Nucleus (Probable).
CC	- TISSUE SPECIFICITY: Expressed in many tissues.
CC	- SIMILARITY: Belongs to the EZ family.
CC	- SIMILARITY: Contains 1 SET domain.
CC	-----
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs license
CC	-----
DR	EMBL; X95653; CAA64955.1; -; mRNA.
DR	EMBL; U61145; AAC51520.1; -; mRNA.
DR	EMBL; U52965; AAC50591.1; -; Genomic_DNA.
DR	PfR; G02838; G02838.
DR	HSSP; Q8X225; IMJ9.
DR	TRANSFAC; T04888; -.
DR	Ensembl; ENSG00000106462; Homo sapiens.
DR	HGNc; HGNC.3527; EZH2.
DR	MIM; 601573; Gene.
DR	GO; GO:0003677; F:DNA binding; TAS.
DR	GO; GO:0006325; P:establishment and/or maintenance of chromatin.; TAS.
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR	InterPro; IPR001005; Myb_DNA_bd.
DR	InterPro; IPR001214; SET_Pfam; PF00856; SET; 1.
DR	SMART; SMART; SAN1.2.
DR	SMART; SMART; SET; 1.
DR	PROSITE; PS50280; SET; 1.
DR	KW Dna-binding; Nuclear protein; Phosphorylation; Transcription;
FW	Transcription regulation.
FT	CHAIN 1 746 Enhancer of zeste homolog 2.
FT	/FTID=PRO_0000213992.
FT	DOMAIN 611 731 SET.
FT	MOTIF 490 495 Nuclear localization signal (Potential).
FT	* COMBIAS 523 605 Cys-rich.
FT	MOD RES 487 487 Phosphothreonine.
FT	CONFLICT 224 224 F -> L (in Ref. 1).
FT	CONFLICT 724 724 F -> V (in Ref. 1).
SQ	SEQUENCE 746 AA; 85363 MW; 1B5029EB9D509BES CRC64;
Query Match	100.0%; Score 4051; DB 1; Length 746;
Best Local Similarity	100.0%; Pred. No. 8.4e-231;
Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1 MGQTGKSKSEKGFCVCRKRYSYMRLQLRRFRADSVSMFSNNOKLERTELINQM 60
DB	1 MGQTGKSKSEKGVCMRKRVKSEYMRLLQLKRRADSVSMFSNNOKLERTELINQM 60
QY	61 KORRIQPVIHLTSVSLSLRTRCSTYSDLPFPQVPLKTLANVAVAPLMYMSPLQNPF 120
DB	61 KORRIQPVHILTSVSLRTRCSTYSDLPFPQVPLKTLANVASVPMTYMSPLQNPF 120
QY	121 MVEDTVALNIIPYMGDEVLDODGTFLBELIKNYDGKVHGDRCGFINDELFVELVNALQ 180
DB	121 MVEDTVALNIIPYMGDEVLDODGTFLBELIKNIDGKNVHGDRCGFINDELFEVLVNALQ 180
QY	181 YNDDDDDDDGDDPEERBEKKOQLEHDRDKESNPERRFPSDKIFEALISMPFDKTAHEL 240
DB	181 YNDDDDDDDGDDPEERBEKKOQLEHDRDKESNPERRFPSDKIFEALISMPFDKTAHEL 240
QY	241 KEKYKELTREQQLPGALPEECTPNIDGPANKSIVOREQSILHSFTLFCRCRCFKYDCFLAPH 300
DB	241 KEKYKELTREQQLPGALPEECTPNIDGPANKSIVOREQSILHSFTLFCRCRCFKYDCFLAPH 300
QY	301 ATPNTYKRNKETALTDNKPCGPCOVCHLGAGBPAAALTAERIKTPPKPCGRRGRRLPN 360
DB	301 ATPNTYKRNKETALTDNKPCGPCOVCHLGAGBPAAALTAERIKTPPKPGRRGRRLPN 360
QY	361 NSSRSTPTINVLIESKDTSDEAGETGEGENNDKEEKKDKETSSSSANSRSCQPIIM 420
DB	361 NSSRSTPTINVLIESKDTSDEAGETGEGENNDKEEKKDKETSSSSANSRSCQPIIM 420
QY	421 KENIIPPENVMSGAEAMFRVLIGTYINDPCAIALIGTKTCROYVEFRVASSIIADPA 480

D	b		421	KPNIEPENVENSGAELSMWRVLTIGTYDNFCAMILGKTKROYEFPVKSSIIAPA	480
Q	y		481	PABVDTPPRKKRKRRLMAHCRKIQLKXDGSSNNHYNTQPCHRPQPCDSSCPCTIAQ	540
D	b		481	PAEDVDPAPKKCKRKRLMAHCRKIQLKXDGSNNHYNYQPCDHPRQPCDSSCPCTIAQ	540
Q	y		541	NFGCKAFQCCSBECONPPRCRCACQCNTHQCPCCTLAVRBCDDPLCLTGAADHMDSKNVS	600
D	b		541	NFGCKAFQCCSECONPPRCRCACQCNTHQCPCCTLAVRBCDDPLCLTGAADHMDSKNVS	600
Q	y		601	CNKCSIORGSKKHLLASPDAVMGFIKPVOVGNKEFTISEYCEIISQDEADRGRKYDX	660
D	b		601	CNKCSIORGSKKHLLASPDAVMGFIKPVOVGNKEFTISEYCEIISQDEADRGRKYDX	660
Q	y		661	YMCSFLENLNDPVADATRKGNKIRPANHSVNENCYAKVMVNGDHRIGIFAPRAIQTGE	720
D	b		661	YMCSFLENLNDFVADATRKGKTLRFANHSVNENPCYAKVMVNGDHRIGIFAPRAIQTGE	720
Q	y		721	ELFPDRYSQADALKYGIRESREMP	746
D	b		721	ELFPDYRYSQADALKYGIRESREMP	746
RESULT 2					
Q	A	R381_MACPA			
ID	A	QAR381_MACPA	PRELIMINARY;	PRT;	746 AA.
AC		QAR381;			
DT	19-JUL-2005,	integrated into UniProtKB/TREMBL.			
DT	19-JUL-2005,	sequence version 1.			
DT	07-FEB-2006,	entry version 5.			
DE	Testis cDNA clone: QcrA-19821,	similar to human enhancer of zeste homolog 2 (Drosophila) (EZHZ), transcript variant 2.			
OS	Macaque fascicularis (Crab eating macaque) (Cynomolgus monkey).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eucarcharia; Euarchontoglires; Primates; Catarrhini;				
OC	Cercopithecoidea; Cercopithecinae; Macaca.				
OX	NCBI_TaxID=9541;				
LN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RX	PubMed=15944441; DOI=10.1093/molbev/msl187;				
RA	Oasada N., Hitaru M., Tanuma R., Kusuda J., Hida M., Suzuki Y., Sugano S., Gojobori T., Shen C.-K.J., Wu C.I., Hashimoto K.;				
RA	"Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis Between Human and Cynomolgus Monkey cDNAs."				
RL	Mol. Biol. Evol. 22:1976-1982(2005).				
LN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RG	International consortium for macaque cDNA sequencing and analysis;				
RT	"DNA sequences of macaque genes expressed in brain or testis and its evolutionary implications."				
RT	submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.				
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms				
CC	Distributed under the Creative Commons Attribution-NoDerivs license				
CC	-----				
CC	EMBL; AB179385; BABO2436.1; -; mRNA.				
DR	GO; GO:0005634; C:nucleus; IEA.				
DR	GO; GO:0003677; F:DNA binding; IEA.				
DR	InterPro; IPR001005; WYB_DNA_bd.				
DR	InterPro; IPR001214; SET.				
DR	Pfam; PF00856; SET; 1.				
DR	SMART; SM00717; SANT; 2.				
DR	SMART; SM00317; SET; 1.				
DR	PROSITE; PS00280; SET; 1.				
SQ	SEQUENCE 746 AA; 8533 MW; 16942199B6E75621 CRC64;				
Query Match 99.9%; Score 4046; DB 2; Length 746;					
Best Local Similarity 99.9%; Pred. No. 1,7e+250;					
Matches 745; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
1 MGCGGGKSGKGPVWKRKVSEYVRLAQLRFRADGVKMFPSNRKILERTLIINQEW 60					



```

Db      1 MGQTGKSEKBPVCMRKVKSEYVRLRLQKFRPRADBYKSMFSSNRQKILERTILNDEM 60
Qy      61 KORRIOPPHILTSVSLRGTRCSTSDLDPTQVTPKLTANAVASVPIYMSWSPLOQNF 120
Db      61 KORRIOPPHILTSVSLRGTRCSTSDLDPTQVTPKLTANAVASVPIYMSWSPLOQNF 120
Qy      121 MVEDETVLNHPYMGDEVLDQGTFFIEBLIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180
Db      121 MVEDETVLNHPYMGDEVLDQGTFFIEBLIKNYDGKVGHDRCGFINDEIFVELVNALGQ 180
Qy      181 YNDDDDDDGDDPBERREKQDLBHRDCKSRPPKPKPSDKIFPAISSMPDKGTABEL 240
Db      181 YNDDDDDDGDDPBERREKQDLBHRDCKSRPPKPKPSDKIFPAISSMPDKGTABEL 240
Qy      241 KEKYKELTEOQLPGALPPECPTNIDGPNAKSVQREOSLSHFTLFCRCFKYDCFLHPH 300
Db      241 KEKYKELTEOQLPGALPPECPTNIDGPNAKSVQREOSLSHFTLFCRCFKYDCFLHPH 300
Qy      301 ATPNTYKRNKTEALDNKPCGPGCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360
Db      301 ATPNTYKRNKTEALDNKPCGPGCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360
Qy      361 NSSRSTPTINVLSEKDTSDREAGTETGENNDKEBEKODTSSSEANRCOTPIKM 420
Db      361 NSSRSTPTINVLSEKDTSDREAGTETGENNDKEBEKODTSSSEANRCOTPIKM 420
Qy      421 KPNIPPEPVENWEGASBAMFRVLIGTYYNFCAIARLIGTKTCROYEPRVKESSIIAPA 480
Db      421 KPNIPPEPVENWEGASBAMFRVLIGTYYNFCAIARLIGTKTCROYEPRVKESSIIAPA 480
Qy      481 PAEDVTPPKRRKRRHRLMAHCRKIQLOKQSSSHVNVNYPCHDPRQCDSSCCCVLAQ 540
Db      481 PAEDVTPPKRRKRRHRLMAHCRKIQLOKQSSSHVNVNYPCHDPRQCDSSCCCVLAQ 540
Qy      541 NFCEKFCOCCSSCCQNRFPQCRCKAOCNTKCCCYLAVERCDPDLCTLGAADHMSKAVS 600
Db      541 NFCEKFCOCCSSCCQNRFPQCRCKAOCNTKCCCYLAVERCDPDLCTLGAADHMSKAVS 600
Qy      601 CNKCSIQGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGEIIISODEADRKKVYDK 660
Db      601 CNKCSIQGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGEIIISODEADRKKVYDK 660
Qy      661 YMCSTFLPMLNDPVVDATRKGNKIRPANHSVNPNCYAKVMVNGDHRIGIPAKRALIQGE 720
Db      661 YMCSTFLPMLNDPVVDATRKGNKIRPANHSVNPNCYAKVMVNGDHRIGIPAKRALIQGE 720
Qy      721 ELFPDYRSQADALKYVGIEREMEIP 746
Db      721 ELFPDYRSQADALKYVGIEREMEIP 746

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Kaha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Mazy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RG NIH MGC Project;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial license
CC -----
DR EMBL, BC010858; AAH10858.1; -; mRNA.
DR HSSP, Q8X225; 1ML9.
DR Ensembl, ENSG00000106462; Homo sapiens.
DR GO, GO:0005634; C:nucleus; IEA.
DR GO, GO:0003677; F:DNA binding; IEA.
DR InterPro, IPR001005; Myb_DNA_bd.
DR InterPro, IPR001214; SET.
DR Pfam, PF00856; SET; 1.
DR SMART, SM00717; SANT; 2.
DR SMART, SM00317; SET; 1.
DR PROSITE, PS00280; SET; 1.
SQ SEQUENCE 751 AA; 86018 MW; D885CF02B60E836 CRC64;

Query Match 99.5%; Score 4030.5; DB 2; Length 751;
Best Local Similarity 99.2%; Pred. No. 1.4e-229;
Matches 745; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

```

QY 476 IIAAPADVDVTPPRKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCHERQPCDSSCP 535  
 DB 481 IIAAPADVDVTPPRKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCHERQPCDSSCP 540  
 QY 536 CVIAKNTCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMD 595  
 DB 541 CVIAKNTCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMD 600  
 QY 596 SKNVSCKNCSIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRG 655  
 DB 601 SKNVSCKNCSIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRG 660  
 QY 656 KVDYKNCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRA 715  
 DB 661 KVDYKNCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRA 720  
 QY 716 IQTGBELFPDYRYSQADALAKYVGIEREWEIP 746  
 DB 721 IQTGBELFPDYRYSQADALAKYVGIEREWEIP 751

RESULT 4  
 Q99L74\_MOUSE PRELIMINARY; PRT; 746 AA.  
 ID Q99L74\_MOUSE  
 AC Q99L74\_MOUSE  
 DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 28.  
 DE Enhancer of zeste homolog 2.  
 GN Name=Enh2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Eumarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
 RC arose spontaneously;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shellen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullanb S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitley M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
 RC arose spontaneously;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
 RC arose spontaneously;  
 RA Director MGC Project;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC -----  
 DR EMBL: BC003772; AAH03772.1; -; mRNA.  
 DR EMBL: BC016391; AAH16391.1; -; mRNA.  
 DR HSSP: Q8X225; IML9.  
 DR EMBL: ENSMUSG0000029687; Mus musculus.  
 DR MGI: MGI:107940; Enh2.  
 DR GO: GO:0045120; C:promoter; IDA.  
 DR GO: GO:0003682; F:chromatin binding; IDA.  
 DR GO: GO:0042054; F:histone methyltransferase activity; IDA.  
 DR GO: GO:0005515; F:protein binding; IPI.  
 DR GO: GO:0016571; P:histone methylation; IDA.  
 DR InterPro: IPR001154; P:negative regulation of striated muscle cell. . .; IDA.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00886; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS0280; SET; 1.  
 SQ SEQUENCE 746 AA; 85292 MW; 7442C751E13EA24B CRC64;

Query Match 98.7%; Score 3999; DB 2; Length 746;  
 Best Local Similarity 98.4%; Pred. No. 9.9e-228;  
 Matches 734; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 MGQTGKSEKGPVCMRKRVSEYRLRLQKFRADSEKMFSSNRKILERTILNQEW 60  
 DB 1 MGQTGKSEKGPVCMRKRVSEYRLRLQKFRADSEKMFSSNRKILERTILNQEW 60  
 QY 61 KQRIQVHILTSYSSIRGREGSVSDLDPPVOVIFLKLTLNVAVPIVYMSPLQNF 120  
 DB 61 KQRIQVHILTSYSSIRGREGSVSDLDPPVOVIFLKLTLNVAVPIVYMSPLQNF 120  
 QY 61 KQRIQVHILTSYSSIRGREGSVSDLDPPVOVIFLKLTLNVAVPIVYMSPLQNF 120  
 DB 61 KQRIQVHILTSYSSIRGREGSVSDLDPPVOVIFLKLTLNVAVPIVYMSPLQNF 120  
 QY 121 MVBDETVLHNPVYMGDEVLDODGFIIEELIKNYGKVHGRECGFINDIFLVELVNLGQ 180  
 DB 121 MVBDETVLHNPVYMGDEVLDODGFIIEELIKNYGKVHGRECGFINDIFLVELVNLGQ 180  
 QY 121 MVBDETVLHNPVYMGDEVLDODGFIIEELIKNYGKVHGRECGFINDIFLVELVNLGQ 180  
 DB 121 MVBDETVLHNPVYMGDEVLDODGFIIEELIKNYGKVHGRECGFINDIFLVELVNLGQ 180  
 QY 181 YNDDDDDDGDDPEREREKQKLEDRHDKESRPPKPFEDKIFEAISMPDDKTAEBL 240  
 DB 181 YNDDDDDDGDDPEREREKQKLEDRHDKESRPPKPFEDKIFEAISMPDDKTAEBL 240  
 QY 181 YNDDDDDDGDDPEREREKQKLEDRHDKESRPPKPFEDKIFEAISMPDDKTAEBL 240  
 DB 181 YNDDDDDDGDDPEREREKQKLEDRHDKESRPPKPFEDKIFEAISMPDDKTAEBL 240  
 QY 241 KEKKEITLQGLPALPPECPTNIDGPNASVQREGLSHSHHTLFCRCRCFYDCELFPH 300  
 DB 241 KEKKEITLQGLPALPPECPTNIDGPNASVQREGLSHSHHTLFCRCRCFYDCELFPH 300  
 QY 241 KEKKEITLQGLPALPPECPTNIDGPNASVQREGLSHSHHTLFCRCRCFYDCELFPH 300  
 DB 241 KEKKEITLQGLPALPPECPTNIDGPNASVQREGLSHSHHTLFCRCRCFYDCELFPH 300  
 QY 301 ATPVYRKQNTETALDNKPCGPCYQHLGAKEPALTLAERIKTPPRPGRRGRGLPN 360  
 DB 301 ATPVYRKQNTETALDNKPCGPCYQHLGAKEPALTLAERIKTPPRPGRRGRGLPN 360  
 QY 301 ATPVYRKQNTETALDNKPCGPCYQHLGAKEPALTLAERIKTPPRPGRRGRGLPN 360  
 DB 301 ATPVYRKQNTETALDNKPCGPCYQHLGAKEPALTLAERIKTPPRPGRRGRGLPN 360  
 QY 361 NSSRSPPTTINVLKSDKTDSDREAGTGTGENNDKEBEKKDETSSSEANSRCQYPIKM 420  
 DB 361 NSSRSPPTTINVLKSDKTDSDREAGTGTGENNDKEBEKKDETSSSEANSRCQYPIKM 420  
 QY 361 NSSRSPPTTINVLKSDKTDSDREAGTGTGENNDKEBEKKDETSSSEANSRCQYPIKM 420  
 DB 361 NSSRSPPTTINVLKSDKTDSDREAGTGTGENNDKEBEKKDETSSSEANSRCQYPIKM 420  
 QY 421 KPNIEPPENVWGAEMSMFVILGYTDNFCALAILGTTGQVYEFKESIIAPA 480  
 DB 421 KPNIEPPENVWGAEMSMFVILGYTDNFCALAILGTTGQVYEFKESIIAPA 480  
 QY 421 KPNIEPPENVWGAEMSMFVILGYTDNFCALAILGTTGQVYEFKESIIAPA 480  
 DB 421 KPNIEPPENVWGAEMSMFVILGYTDNFCALAILGTTGQVYEFKESIIAPA 480  
 QY 481 PAEDVDPKPKKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCDHPQPCDSSCPVIAQ 540  
 DB 481 PTEBDVDPKPKKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCDHPQPCDSSCPVIAQ 540  
 QY 481 PTEBDVDPKPKKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCDHPQPCDSSCPVIAQ 540  
 DB 481 PTEBDVDPKPKKKKKRHLMAHCRKIQLKKGSSNNHYNYQPCDHPQPCDSSCPVIAQ 540  
 QY 541 NFCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMDSKNVS 600  
 DB 541 NFCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMDSKNVS 600  
 QY 541 NFCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMDSKNVS 600  
 DB 541 NFCEKFCQSSSECONFPFGCRCAQNTKQPCYLAVERCDPLCTCGAADHMDSKNVS 600  
 QY 601 CKNCISIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRGKYYDK 660  
 DB 601 CKNCISIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRGKYYDK 660  
 QY 601 CKNCISIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRGKYYDK 660  
 DB 601 CKNCISIORGSKHLLAPSDVAGMGIFIDPVQKNEFISEYCGEITISODEADRGKYYDK 660  
 QY 661 YMCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRAIQTE 720  
 DB 661 YMCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRAIQTE 720  
 QY 661 YMCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRAIQTE 720  
 DB 661 YMCSPFLFNANDPVVDATKGNKIRFANHSVNPNCYAKVMVNGDRHIGIPAKRAIQTE 720

Db 661 YMCSEFLFNANDPVVDATRKNGKIRFANSHVNPNCYAKVMVNGDHRIGIPAKRAIQTGE 720

Qy 721 ELFPDPRYSQADALAKTVGIEREMETIP 746

Db 721 ELFPDPRYSQADALAKTVGIEREMETIP 746

RESULT 5

OS71L5\_MOUSE PRELIMINARY; PRT; 779 AA.

ID OS71L5\_MOUSE

AC 0571L5; 10-MAY-2005, integrated into UniProtKB/TrEMBL.

DT 10-MAY-2005, sequence version 1.

DT 07-FEB-2006, entry version 4.

DE MKIAA4065 protein (fragment).

GN Name=Ezh2; Synonyms=MKIAA4065;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

OC NCBI\_Taxid=10090;

OK NCBI\_Taxid=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=embryonic tail;

RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hiraoka S., Suga Y., Nagase T., Ohara O., Koga H.;

RT Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs Identified by Screening of Terminal sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries. "

RT Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.

RL

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonCommercial license

CC -----

DR EMBL; AK220174; BAD90359.1; -; mRNA.

DR Ensembl; ENSMUSG0000029687; Mus musculus.

DR MGI; MGI:107940; Ezh2.

DR GO; GO:0045120; C:chromatin; IDA.

DR GO; GO:0003682; F:chromatin binding; IDA.

DR GO; GO:0042054; F:histone methyltransferase activity; IDA.

DR GO; GO:0005515; F:protein binding; IPT.

DR GO; GO:0016571; P:histone methylation; IDA.

DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.

DR InterPro; IPR001005; MYD\_DNA\_Bd.

DR InterPro; IPR001214; SET.

DR Pfam; PF00856; SET; 1.

DR SMART; SM00317; SANT; 2.

DR SMART; SM00317; SET; 1.

DR PROSITE; PS50280; SET; 1.

FT NON TER 1

FT SEQUENCE 779 AA; 88523 MW; A0CC1A16CA868702 CRC64;

Query Match 98.7%; Score 3999; DB 2; Length 779;

Best Local Similarity 98.4%; Pred. No. 1e-227;

Matches 734; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MGQTGKSKBGPVCRKVKSEYVRLRQLKRRFRADEVKSMFSSNRKILERTETIANQW 60

Db 34 MGQTGKSKBGPVCRKVKSEYVRLRQLKRRFRADEVKSMFSSNRKILERTETIANQW 93

Qy 61 KORRIQPHILTSVSLAGTREGCTYSLDLPPTQYIPLKLTANAASVPIKMSWSPLOQNF 120

Db 94 KORRIQPHIMTSVSLGTRCSTYSDLPFAQVIPLKLTANAASVPIKMSWSPLOQNF 153

Qy 121 MVEDEVVLAHNTPYMDDEVLDODGTIFBELIKKYDGKVGDRCEGFINDEIFVELNALGQ 180

Db 154 MVEDEVVLAHNTPYMDDEVLDODGTIFBELIKKYDGKVGDRCEGFINDEIFVELNALGQ 213

Qy 181 YNDDDDDDGDDPDEERBEKQDLBDHARDKESRPKRKPSDKIFPAISMPEDKTAEBL 240

Db 214 YNDDDDDDGDDPDEERBEKQDLBDHARDKESRPKRKPSDKIFPAISMPEDKTAEBL 273

Qy 241 KEKYBELTEQOLPGALPPECTENIDGPNKASVQREQSILHSFHTLFCRRCFKYDCFLHPH 300

Db 274 KEKYBELTEQOLPGALPPECTENIDGPNKASVQREQSILHSFHTLFCRRCFKYDCFLHPH 333

Qy 301 ATPNTYKKNKTETALDNKRCGPQCYOHLGAKAFPAALTAERIKTTPKRPGRRRGRLPN 360

Db 334 ATPNTYKKNKTETALDNKRCGPQCYOHLGAKAFPAALTAERIKTTPKRPGRRRGRLPN 393

Qy 361 NSSRSTPTINVLSEKOTDSREAGTETGGENNDGEBBKKDETSSSSANRCOTPIKM 420

Db 394 NSSRSTPTINVLSEKOTDSREAGTETGGENNDGEBBKKDETSSSSANRCOTPIKM 453

Qy 421 KPNIPPEVWESGAASMFRLIGTYTDNFCALRLIGTKRCROYEPRVSSIIAPA 480

Db 454 KPNIPPEVWESGAASMFRLIGTYTDNFCALRLIGTKRCROYEPRVSSIIAPA 513

Qy 481 PAEDVTPPRKKRKRRLMAAHCRIQLKQDSSNHVNYOPCDHPRQPCDSSCPVIAQ 540

Db 514 PTEDVTPPRKKRKRRLMAAHCRIQLKQDSSNHVNYOPCDHPRQPCDSSCPVIAQ 573

Qy 541 NFCEKPCQCCSQCQRNPFQCRCKAQCNTRKQPCYLAVERCDPLCLTCGADHMSKYN 600

Db 574 NFCEKPCQCCSQCQRNPFQCRCKAQCNTRKQPCYLAVERCDPLCLTCGADHMSKYN 633

Qy 601 CKNCSIQGSKKHLIAPSDVAGWIFIKDPVQKNEFISEYCGEIIISODEADRGKYDX 660

Db 634 CKNCSIQGSKKHLIAPSDVAGWIFIKDPVQKNEFISEYCGEIIISODEADRGKYDX 693

Qy 661 YMCSEFLFNANDPVVDATRKNGKIRFANSHVNPNCYAKVMVNGDHRIGIPAKRAIQTGE 720

Db 694 YMCSEFLFNANDPVVDATRKNGKIRFANSHVNPNCYAKVMVNGDHRIGIPAKRAIQTGE 753

Qy 721 ELFPDPRYSQADALAKTVGIEREMETIP 746

Db 754 ELFPDPRYSQADALAKTVGIEREMETIP 779

RESULT 6

EZH2\_MOUSE STANDARD; PRT; 746 AA.

ID EZH2\_MOUSE

AC Q61188; Q9R090; 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.

DT 01-NOV-1996, sequence version 1.

DT 07-FEB-2006, entry version 48.

DE Enhancer of zeste homolog 2 (ENX-1).

DS Enhancer of zeste homolog 2 (ENX-1).

OS Mus musculus (Mouse).

GN Name=Ezh2; Synonyms=Enx1h;

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

OC NCBI\_Taxid=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Placenta;

RA MEDLINE=97014262; Pubmed=8861097; DOI=10.1016/0925-4773(96)00499-6; Hobert O., Sures I., Closek T., Fuchs M., Ullrich A.;

RT "Isolation and developmental expression analysis of Enx-1, a novel mouse Polycomb group gene."

RT Mech. Dev. 55:171-184(1996).

RL

RN [2]

RP NUCLEOTIDE SEQUENCE OF 134-497.

RC STRAIN=128/SV;

RA MEDLINE=99160476; Pubmed=10051331; DOI=10.1007/s003359900993; Leible G., Haynes A.R., Lebersorger A., O'Carroll D., Mattei M.-G., Denny P., Brown S.D.M., Jenuwein T.;

RT "The murine polycomb-group genes ezh1 and ezh2 map close to hox gene clusters on mouse chromosomes 11 and 6."

RL Mamm. Genome 10:311-314(1999).

CC -1- FUNCTION: May be involved in the regulation of gene transcription and chromatin structure.

CC -1- SUBUNIT: Binds ATRX via the SET domain (By similarity).

CC -1- SUBCELLULAR LOCATION: Nucleus (Probable).

CC -1- ALTERNATIVE PRODUCTS:

```
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=ENX-1A;
CC IsoId=061188-1; Sequence=Displayed;
CC Name=ENX-1B;
CC IsoId=061188-2; Sequence=VSP_001501;
CC -1- TISSUE SPECIFICITY: Widely expressed in early embryos. In later
CC embryogenesis, expression restricted to central and peripheral
CC nervous system, liver and thymus. In adult, highest expression in
CC spleen, testis and placenta. Lower levels in intestine and muscle
CC and very low levels in brain and liver. No expression in heart,
CC thyroid gland, lung and kidney.
CC -1- DEVELOPMENTAL STAGE: Expressed in both adult and embryo with
CC highest levels in early embryogenesis.
CC -1- SIMILARITY: Belongs to the E2 family.
CC -1- SIMILARITY: Contains 1 SET domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC -----
DR EMBL; U52951; AAC5265.1; -; mRNA.
DR EMBL; AF104359; AAD54020.1; -; Genomic DNA.
DR HSSP; O8X225; 1ML9.
DR Ensembl; ENSMUSG00000029687; Mus musculus.
DR MGI; MGI:107940; E2n2.
DR LinkHub; Q61188; -.
DR GO; GO:0045120; C:pronucleus; IDA.
DR GO; GO:0003264; F:chromatin binding; IDA.
DR GO; GO:0042062; F:histone methyltransferase activity; IDA.
DR GO; GO:0005515; P:protein binding; IPI.
DR GO; GO:0016571; P:histone methylation; IDA.
DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.
DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00317; SANT; 2.
DR PROSITE; PS50280; SET; 1.
DR Alternative splicing; DNA-binding; Nuclear protein; Phosphorylation;
DR Transcription; Transcription regulation.
DR CHAIN 1 746 Enhancer of zeste homolog 2.
FT DOMAIN 611 731 SET.
FT MOTIF 490 495 Nuclear localization signal (Potential).
FT COMBIS 523 605 Cys-rich.
FT MOD_RES 487 487 Phosphothreonine (By similarity).
FT VARSPLIC 511 553 DGSNNHYNQCDDPDRPCDSCPCVLAQNFCEKFCQCSS
FT EC -> G (in isoform ENX-1B).
FT CONFLICT 159 161 /FTID=VSP_001501.
FT SEQ 746 AA; 85336 MW; 0435C021963BD24E CRC64;
SQ SEQUENCE 746 AA; 85336 MW; 0435C021963BD24E CRC64;
Query Match 98.6%; Score 3993; DB 1; Length 746;
Best Local Similarity 98.3%; Pred. No. 2.2e-227;
Matches 733; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
```

```
DB 241 KKKYKLTBOQLPALPPECTPNIIDGNASVQREGSLHSFTLLFCRCRCFKYDCFLHFFH 300
QY 301 ATPYTYRKNKTETALDNKPCGPOCYOHLGAKGFAALTAERIKTPPKRPGRRRGLPN 360
DB 301 ATPYTYRKNKTETALDNKPCGPOCYOHLGAKGFAALTAERIKTPPKRPGRRRGLPN 360
QY 361 NSRSPSTPTTNVLESKOTBDBRAGTGTGENDDKBEKKDFTSSSSSEANSQCPTIYM 420
DB 361 NSRSPSTPTTNVLESKOTBDBRAGTGTGENDDKBEKKDFTSSSSSEANSQCPTIYM 420
QY 421 KPNTPEPENTWMSGASBMPVLIIGTYDNPFAIARLIGTKGRQYEPFVKESSIIAPA 480
DB 421 KPNTPEPENTWMSGASBMPVLIIGTYDNPFAIARLIGTKGRQYEPFVKESSIIAPA 480
QY 481 PAEDVDTPPKRKRKRLMAHCRKIQLKDGSSNHYNYNQCDDPDRPCDSSCPCYIAQ 540
DB 481 PTEDVDTPPKRKRKRLMAHCRKIQLKDGSSNHYNYNQCDDPDRPCDSSCPCYIAQ 540
QY 541 NPKRPFQCGSSSECONRPPGRCRAQCCTKQCPCYLAVERCDPDLCTCGAADHWSKNVS 600
DB 541 NPKRPFQCGSSSECONRPPGRCRAQCCTKQCPCYLAVERCDPDLCTCGAADHWSKNVS 600
QY 601 CRKCSIORGSKKHLAPDVAAGWGFITKDPVQKNRFTSYCEBIIISQDEADRRGKYDK 660
DB 601 CRKCSIORGSKKHLAPDVAAGWGFITKDPVQKNRFTSYCEBIIISQDEADRRGKYDK 660
QY 661 YMCSPFLNLNNDPVADTRKGNKIRFANHSVNPVCAYVMVNGDRHIGIPAKRAIOTGE 720
DB 661 YMCSPFLNLNNDPVADTRKGNKIRFANHSVNPVCAYVMVNGDRHIGIPAKRAIOTGE 720
QY 721 ELFPDYYRQADALKYGIEREWEIP 746
DB 721 ELFPDYYRQADALKYGIEREWEIP 746
RESULT 7
Q37ZH6_MOUSE PRELIMINARY; PRT; 746 AA.
AC Q37ZH6;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Erythroblast cDNA, RIKEN full-length enriched library,
DE clone:K0C026B09 product:enhancer of zeste homolog 2 (Drosophila),
DE full insert sequence.
OS Name=Bzh2;
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STPAIN=C57BL/6J;
RX PubMed=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RT Carninci P., Hayashizaki Y.;
RL "High-efficiency full-length cDNA cloning.";
Methods Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STPAIN=C57BL/6J;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kaasakawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilmong L.G., Aldinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Banal M., Baxter L., Beisel K.W., Beresno T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christofels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.B.,
```

RA Guetlich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Humnicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Ueki M., Kanapin A., Katon M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Larcrau L.F., Lazarovic D., Lipovich L., Liu J.,  
RA Liuni S., McMillan S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Maruzawa S., Miki H., Mignone P., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavessi G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Roet B., Ruan Y., Salberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Sempke C.A., Sero S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugliura K., Sultana R., Takenaka Y., Taki K.,  
RA Tamajo K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yaqi K.,  
RA Yamanishi H., Zharovskiy E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brucic V., Quackenbush J.,  
RA Walsted C., Wattick J.S., Hume D.A., Kai C., Saeki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida Y., Imanishi K., Itoh M., Kato T., Kawaji H., Kawagashita N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Wataniki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.,  
RT "The transcriptomic landscape of the mammalian genome.",  
RL Science 309:1559-1563(2005).  
[3]  
RA NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RT (Genome Network Core Team) and the FANTOM Consortium;  
RL "Antisense Transcription in the Mammalian Transcriptome.",  
RN Science 309:1564-1566(2005).  
[4]  
RA NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RG Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nkado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Betzel K.W.,  
RA Blake J.A., Bratt D., Brusic J., Chothia C., Corbani L.B., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimonod S., Guetlich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedziński R.M., King B.L.,  
RA Kanagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Malais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Petrea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyman-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalina M., Zhu Y., Zimmer A., Carlini P., Hayatsu N.,  
RA Hirozane-Koriyama T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Haseizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yamanishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.",  
RL Nature 420:563-573(2002).  
[5]  
RA NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
[6]

RA Kawai J, Shihagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y,  
RA Arkawa T, Hara A, Fukunishi Y, Komoto H, Adachi J, Fukuda S,  
RA Arkawa K, Iwawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaoka I,  
RA Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R,  
RA Kadota K, Matsuda H.A., Ashburner M, Batalov S, Casavant T,  
RA Flischiemann W, Gaasterland T, Gissi C, King B, Kochia H.,  
RA Knuehl P, Lewis S, Matsuo Y, Nikaido I., Pesele G, Quackenbush J,  
RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barth G.,  
RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guetioncin S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welts C., Whitaker C., Wilmink L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohatsu S.,  
RA Hayashizaki Y. ;  
RT "Functional annotation of a full-length mouse cDNA collection." ;  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Komoto H., Okazaki Y., Muramatsu M., Hayashizaki Y. ;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RL prepare full-length cDNA libraries for rapid discovery of new genes." ;  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=20530913; PubMed=11047861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,  
RA Komoto H., Akiyama J., Nishi K., Kitamura T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishize T., Harada A.,  
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y. ;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RL sequencing pipeline with 384 multicapillary sequencer." ;  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RA Arkawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hoti F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Komoto H., Murota M., Nakamura M., Nitomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watabiki A.,  
RA Muramatsu M., Hayashizaki Y. ;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC  
CC EMBL; AK157953; BAR34232.1; -; mRNA.  
DR MGI; MGI:107940; Ezh2.  
DR GO; GO:0045120; C:chromatin binding, IDA.  
DR GO; GO:0003682; F:chromatin binding, IDA.  
DR GO; GO:0042054; F:histone methyltransferase activity, IDA.  
DR GO; GO:0005515; F:protein binding, IPI.  
DR GO; GO:0016571; P:histone methylation, IDA.  
DR GO; GO:0051154; P:negative regulation of striated muscle cell. . . , IDA  
DR InterPro; IPR001005; MYD\_DNA\_Bd.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET, 1.  
DR SMART; SM00717; SANT, 2.  
DR SMART; SM00317; SET, 1.  
DR PROSITE; PSS0280; SET, 1.  
SQ SEQUENCE 746 AA, 85332 MW, BAD303CCD9CB6FA4 CRC64;

Query Match 98.5%; Score 3990; DB 2; Length 746;  
 Best Local Similarity 98.3%; Pred. No. 3.3e-227;  
 Matches 733; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 MGOTGKSEKGPVCMRKRVSEYWRQLQKFRRADEVKSMFSSNRKILERTITLQEW 60  
 DB 1 MGOTGKSEKGPVCMRKRVSEYWRQLQKFRRADEVKSMFSSNRKILERTITLQEW 60  
 QY 61 KORIQPHILTSVSLRGTRCSVSDLPPTVPIPLKTLNAVASVIMYSPLQONF 120  
 DB 61 KORIQPHILTSVSLRGTRCSVSDLPPTVPIPLKTLNAVASVIMYSPLQONF 120  
 QY 121 MVEDEVLANIPYMGDEVLDQDGFIFIELIKNYDGKVGHRECGFINDEIFVELVNALGQ 180  
 DB 121 MVEDEVLANIPYMGDEVLDQDGFIFIELIKNYDGKVGHRECGFINDEIFVELVNALGQ 180  
 QY 181 YNDDDDDDGDDPEREREKQDLBDHDKESRPPRPSPKTIPEALISSMPDQGTABEL 240  
 DB 181 YNDDDDDDGDDPEREREKQDLBDHDKESRPPRPSPKTIPEALISSMPDQGTABEL 240  
 QY 241 KEKTKELTEQOLPGALPPECTPNIDGNPAKSVOREGSLHSFHTLFCRRCFYDCFLHPFH 300  
 DB 241 KEKTKELTEQOLPGALPPECTPNIDGNPAKSVOREGSLHSFHTLFCRRCFYDCFLHPFH 300  
 QY 301 ATPPTYRKNTETALDNKPCGPOCYOHLBGAKFPAALTAERIKTPPRPGRRRLPN 360  
 DB 301 ATPPTYRKNTETALDNKPCGPOCYOHLBGAKFPAALTAERIKTPPRPGRRRLPN 360  
 QY 361 NSSRSPPTINVLKSTQDSREAGTETGEMNKKEBEKDETSSEANSRCOTPLKM 420  
 DB 361 NSSRSPPTINVLKSTQDSREAGTETGEMNKKEBEKDETSSEANSRCOTPLKM 420  
 QY 421 KPNIPEPENVMGSAEMFVILGTYDNCALIALIGTKYCRQVYEFYKESIIAPA 480  
 DB 421 KPNIPEPENVMGSAEMFVILGTYDNCALIALIGTKYCRQVYEFYKESIIAPA 480  
 QY 481 PAEDVDTPPRKKRHRMAHCKRIQLKKGSSNNHYNYOPCDHPROPCDSCPCVIAQ 540  
 DB 481 PAEDVDTPPRKKRHRMAHCKRIQLKKGSSNNHYNYOPCDHPROPCDSCPCVIAQ 540  
 QY 541 NPCKFCQCCSECONRPPGCRCKAOCNKCPCVIAVECPDCLTCGADHMDSKNVS 600  
 DB 541 NPCKFCQCCSECONRPPGCRCKAOCNKCPCVIAVECPDCLTCGADHMDSKNVS 600  
 QY 601 CKNCISIGSKKHLILAPSDVAGWGIPIKDPVQKNEFISBYCGRIISQDEADRRGKYDK 660  
 DB 601 CKNCISIGSKKHLILAPSDVAGWGIPIKDPVQKNEFISBYCGRIISQDEADRRGKYDK 660  
 QY 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKALQTGE 720  
 DB 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKALQTGE 720  
 QY 721 ELFPDYRSQADALKYVGIEREMEIP 746  
 DB 721 ELFPDYRSQADALKYVGIEREMEIP 746

RESULT 8  
 06AXH7\_MOUSE PRELIMINARY; PRT; 742 AA.  
 AC 06AXH7;  
 DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
 DT 13-SEP-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 15.  
 DE Ezh2 protein.  
 GN Name=Ezh2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=2368257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA DiCicco L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Mada A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Mada A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buttefield Y.S.N., Krzywicki M.I., Skalka U., Smallus D.E.,  
 RA Schermer A., Schein J.B., Jones S.U.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Director MGC Project;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivative License  
 CC -----  
 DR EMBL; BC079538; AAH79538.1; -; mRNA.  
 DR Ensembl; ENSMUSG00000029687; Mus musculus.  
 DR MGI; MGI:107940; Ezh2.  
 DR GO; GO:0045120; C:promoter; IDA.  
 DR GO; GO:0003682; P:chromatin binding; IDA.  
 DR GO; GO:0042054; F:histone methyltransferase activity; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0016571; P:histone methylation; IDA.  
 DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.  
 DR InterPro; IPR001005; Myb\_DNA\_bd.  
 DR InterPro; IPR001214; SET.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00717; SANT; 2.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PSS0280; SET; 1.  
 SO SEQUENCE 742 AA; 84934 MW; A9CDB80ACB2C6D6A CRC64;

Query Match 96.8%; Score 3922; DB 2; Length 742;  
 Best Local Similarity 96.4%; Pred. No. 3.5e-223;  
 Matches 724; Conservative 7; Mismatches 6; Indels 14; Gaps 2;

QY 1 MGOTGKSEKGPVCMRKRVSEYWRQLQKFRRADEVKSMFSSNRKILERTITLQEW 60  
 DB 1 MGOTGKSEKGPVCMRKRVSEYWRQLQKFRRADEVKSMFSSNRKILERTITLQEW 60  
 QY 61 KORIQPHILTSVSLRGTRCSVSDLPPTVPIPLKTLNAVASVIMYSPLQONF 120  
 DB 61 KORIQPHILTSVSLRGTRCSVSDLPPTVPIPLKTLNAVASVIMYSPLQONF 120  
 QY 121 MVEDEVLANIPYMGDEVLDQDGFIFIELIKNYDGKVGHRECGFINDEIFVELVNALGQ 180  
 DB 121 MVEDEVLANIPYMGDEVLDQDGFIFIELIKNYDGKVGHRECGFINDEIFVELVNALGQ 180  
 QY 181 YNDDDDDDGDDPEREREKQDLBDHDKESRPPRPSPKTIPEALISSMPDQGTABEL 240  
 DB 181 YNDDDDDDGDDPEREREKQDLBDHDKESRPPRPSPKTIPEALISSMPDQGTABEL 240  
 QY 241 KEKTKELTEQOLPGALPPECTPNIDGNPAKSVOREGSLHSFHTLFCRRCFYDCFLHPFH 300  
 DB 241 KEKTKELTEQOLPGALPPECTPNIDGNPAKSVOREGSLHSFHTLFCRRCFYDCFLHPFH 300  
 QY 301 ATPPTYRKNTETALDNKPCGPOCYOHLBGAKFPAALTAERIKTPPRPGRRRLPN 360  
 DB 301 ATPPTYRKNTETALDNKPCGPOCYOHLBGAKFPAALTAERIKTPPRPGRRRLPN 360  
 QY 361 NSSRSPPTINVLKSTQDSREAGTETGEMNKKEBEKDETSSEANSRCOTPLKM 420  
 DB 361 NSSRSPPTINVLKSTQDSREAGTETGEMNKKEBEKDETSSEANSRCOTPLKM 420  
 QY 421 KPNIPEPENVMGSAEMFVILGTYDNCALIALIGTKYCRQVYEFYKESIIAPA 480  
 DB 421 KPNIPEPENVMGSAEMFVILGTYDNCALIALIGTKYCRQVYEFYKESIIAPA 480  
 QY 481 PAEDVDTPPRKKRHRMAHCKRIQLKKGSSNNHYNYOPCDHPROPCDSCPCVIAQ 540  
 DB 481 PAEDVDTPPRKKRHRMAHCKRIQLKKGSSNNHYNYOPCDHPROPCDSCPCVIAQ 540  
 QY 541 NPCKFCQCCSECONRPPGCRCKAOCNKCPCVIAVECPDCLTCGADHMDSKNVS 600  
 DB 541 NPCKFCQCCSECONRPPGCRCKAOCNKCPCVIAVECPDCLTCGADHMDSKNVS 600  
 QY 601 CKNCISIGSKKHLILAPSDVAGWGIPIKDPVQKNEFISBYCGRIISQDEADRRGKYDK 660  
 DB 601 CKNCISIGSKKHLILAPSDVAGWGIPIKDPVQKNEFISBYCGRIISQDEADRRGKYDK 660  
 QY 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKALQTGE 720  
 DB 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKALQTGE 720  
 QY 721 ELFPDYRSQADALKYVGIEREMEIP 746  
 DB 721 ELFPDYRSQADALKYVGIEREMEIP 746

DB 292 SYSFATINTYKRTETALDNKPCGPCYOHLEGAKEFAAALTHERIKTPKRCGRR 351  
QY 356 GRLPNNRRPSTPTINVLSEKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQ 415  
DB 352 GRLPNNRRPSTPTISVLESKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQ 411  
QY 416 TPRIKKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSS 475  
DB 412 TPRIKKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSS 471  
QY 476 IIPAPADVDTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCP 535  
DB 472 IIPAPADVDTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCP 531  
QY 536 CVIAQNFCEKPCQSSSECONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 595  
DB 532 CVIAQNFCEKPCQSSSECONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 591  
QY 596 SKNVCCKRCSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 655  
DB 592 SKNVCCKRCSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 651  
QY 656 KYVDKYMCSFLPNLNDPVVDATRKGNKLRPANHSVNPNCYAKVMVNGDRHIGIPAKRA 715  
DB 652 KYVDKYMCSFLPNLNDPVVDATRKGNKLRPANHSVNPNCYAKVMVNGDRHIGIPAKRA 711  
QY 716 IQTGEELFPDYRYSGADALKTGYIEREMEIP 746  
DB 712 IQTGEELFPDYRYSGADALKTGYIEREMEIP 742

RESULT 9  
Q5XH80 XENLA PRELIMINARY; PRT; 748 AA.

AC Q5XH80: 23-NOV-2004, integrated into UniProtKB/TrEMBL.  
DT 23-NOV-2004, sequence version 1.  
DT 21-FEB-2006, entry version 15.  
DE E2h2 protein.  
GN Name=E2h2;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Butelaeostomi;  
CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
NCBI TaxID=8355;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Ovary;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative."  
RT Dev. Dyn. 225:364-391 (2002).  
RL [2]  
RN NUCLEOTIDE SEQUENCE.  
RC TISSUE=Ovary;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheet C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marulanda K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stedman M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Foley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Scherich A., Schein J.E., Jones S.J.M., Maira M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Ovary;  
RA Klein S., Gerthard D.S.;  
RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs license  
CC -----  
DR EMBL; BC084193; AAH84193.1; -; mRNA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR001005; Myb\_DNA\_bd.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00317; SANT; 2.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS0280; SET; 1.  
SQ SEQUENCE 748 AA; 8555 MW; 2916EA4B3CB8C4E CRC64;  
  
Query Match 94.0%; Score 3809; DB 2; Length 748;  
Best Local Similarity 93.4%; Pred. No. 1.6e-216;  
Matches 699; Conservative 25; Mismatches 22; Indels 2; Gaps 1;  
  
QY 1 MGOTGKSKGKGVCMWRKRVSEVMRLQKRRPADVYKSSNRKQILERTILN 60  
DB 1 MGOTGKSKGKGVCMWRKRVSEVMRLQKRRPADVYKSSNRKQILERTILN 60  
QY 61 KORIOPVHILTVSSLRGTGECVSTDLDPPTQVPIPLKTLNAAVAPIMYSWPLQONF 120  
DB 61 KORIOPVHILTVSSLRGTGECVSTDLDPPTQVPIPLKTLNAAVAPIMYSWPLQONF 120  
QY 121 MVEDETVLANIPYMGDEVLDQDGFIEELIKNYDGKVGHDEGCFINDEIFELVNALGQ 180  
DB 121 MVEDETVLANIPYMGDEVLDQDGFIEELIKNYDGKVGHDEGCFINDEIFELVNALGQ 180  
QY 121 MVEDETVLANIPYMGDEVLDQDGFIEELIKNYDGKVGHDEGCFINDEIFELVNALGQ 180  
DB 121 MVEDETVLANIPYMGDEVLDQDGFIEELIKNYDGKVGHDEGCFINDEIFELVNALGQ 180  
QY 181 YNDDDDDDDGDD--PERRERKQDLBHRDKESRPRKPSDKIFRALSMPDDKGA 238  
DB 181 YNDDDDDDDGDD--PERRERKQDLBHRDKESRPRKPSDKIFRALSMPDDKGA 238  
QY 181 YSDYEDDEDGDDDDQDDTAKQDDNMDKQPLRKPSSDKIFRALSMPDDKGS 240  
DB 181 YSDYEDDEDGDDDDQDDTAKQDDNMDKQPLRKPSSDKIFRALSMPDDKGS 240  
QY 239 ELKRYKELTEQQLPGALPPECTPNIIDGPNKASYQRQSLHSFTLLFCRRCFYKDFLHP 298  
DB 239 ELKRYKELTEQQLPGALPPECTPNIIDGPNKASYQRQSLHSFTLLFCRRCFYKDFLHP 298  
QY 241 ELKRYKELTEQQLPGALPPECTPNIIDGPNKASYQRQSLHSFTLLFCRRCFYKDFLHP 300  
DB 241 ELKRYKELTEQQLPGALPPECTPNIIDGPNKASYQRQSLHSFTLLFCRRCFYKDFLHP 300  
QY 299 FHATPNTYKRRNTETALDNKPCGPCYOHLEGAKEFAAALTHERIKTPPKRPGRRRRL 358  
DB 299 FHATPNTYKRRNTETALDNKPCGPCYOHLEGAKEFAAALTHERIKTPPKRPGRRRRL 358  
QY 301 FHATPNTYKRRNTETALDNKPCGPCYOHLEGAKEFAAALTHERIKTPPKRPGRRRRL 360  
DB 301 FHATPNTYKRRNTETALDNKPCGPCYOHLEGAKEFAAALTHERIKTPPKRPGRRRRL 360  
QY 359 PNNSSRPSTPTINVLSEKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQPTI 418  
DB 359 PNNSSRPSTPTINVLSEKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQPTI 418  
QY 361 PNNSSRPSTPTINVLSEKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQPTI 420  
DB 361 PNNSSRPSTPTINVLSEKOTSDREAGTETGENDKEERKKDRTSSSSSANSRCQPTI 420  
QY 419 KMKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSSIIA 478  
DB 419 KMKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSSIIA 478  
QY 421 KMKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSSIIIS 480  
DB 421 KMKPNIEPPEVMSGASBMSFRVLIGTYDNPFCALRIIGTKTCROYEFVEXSSIIIS 480  
QY 479 PAPAADVTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCPV 538  
DB 479 PAPAADVTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCPV 538  
QY 481 PAPAADVTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCPV 540  
DB 481 PAPAADVTPPRKKRKRHLMAHCRKIOLKXGSSNNHYNYOPCDHPKPCDSSCPV 540  
QY 539 AONPCERFCCSSSCONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 598  
DB 539 AONPCERFCCSSSCONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 598  
QY 541 AONPCERFCCSSSCONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 600  
DB 541 AONPCERFCCSSSCONRPFGRCAQCNTPKOCPCYLAVERCDPDLCTCGAADM 600  
QY 599 VSCNKGSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 658  
DB 599 VSCNKGSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 658  
QY 601 VSCNKGSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 660  
DB 601 VSCNKGSIORGSKKHLAPSPVAGMGIPIKDPVQKNEFTISYCGEIIISQDEADR 660



```

QY 659 DKTWCSEFLFNILNDFVVDATRKGNKIRFANHNVNPNVCYAKVMVNGDHRIGIFAKRAIQT 718
DB 661 DKTCMSFLFNILNDFVVDATRKGNKIRFANHNVNPNVCYAKVMVNGDHRIGIFAKRAIQT 720
QY 719 GEBLFPDYRYSQADALKTGVIEREMEIP 746
DB 721 GEBLFPDYRYSQADALKTGVIEREMEIP 748

RESULT 10
Q98SM3_XENLA PRELIMINARY; PRT; 748 AA.
ID Q98SM3_XENLA
AC Q98SM3;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Enhancer of zeste.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21184113; PubMed=11287189; DOI=10.1016/S0925-4773(01)00304-5;
RA Barnett M.W., Seville R.A., Nijjar S., Old R.W., Jones B.A.;
RT "Xenopus Enhancer of Zeste (XZ): an anteriorly restricted polycomb
RL gene with a role in neural patterning."
RL Mech. Dev. 102:157-167(2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC -----
DB EMBL: AF351126; AAK30208.1; -, mRNA.
DR HSSP; Q8X225; IML9.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001005; MYD_DNA_bd.
DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00717; SANT; 2.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50280; SET; 1.
SQ SEQUENCE 748 AA; 85383 MW; 8618F615B348C46C CRC64;

Query Match 93.6%; Score 3793; DB 2; Length 748;
Best Local Similarity 93.0%; Pred. No. 1.4e-215;
Matches 696; Conservative 26; Mismatches 24; Indels 2; Gaps 1;

QY 1 MGQTGKSEKGPVCMRKRVKSEYMRQLQKRRADDEVKSMFSSNRQKILERTILNQEW 60
DB 1 MGQTGKSEKGPVCMRKRVKSEYMRQLQKRRADDEVKSMFSSNRQKILERTILNQEW 60
QY 61 KORIOPHILITSVSLGTRGCVTSLDLPPTQYIPLKTNANAVASVPIMTSKPLQONF 120
DB 61 KORIOPHILITSVSLGTRGCVTSLDLPPTQYIPLKTNANAVASVPIMTSKPLQONF 120
QY 121 MVBETVLAMNIPVNGDEVLDQGTFFIELIKNYGKVGNDRCGFINDEIVELVNALGQ 180
DB 121 MVBETVLAMNIPVNGDEVLDQGTFFIELIKNYGKVGNDRCGFINDEIVELVNALGQ 180
QY 181 YNDDDDDDDGDD--PERRERKQDLERHDDKESRPPKFPSPDKI FFAISSMPFDKGTAE 238
DB 181 YNDDDDDDDGDD--PERRERKQDLERHDDKESRPPKFPSPDKI FFAISSMPFDKGTAE 238
QY 239 ELKERYKELTQQLPGALPPECTPNIDGPNAKSVOREGSLSPFTLFCRRKPKTDCFLHP 298
DB 239 ELKERYKELTQQLPGALPPECTPNIDGPNAKSVOREGSLSPFTLFCRRKPKTDCFLHP 298
QY 241 ELKERYKELTQQLPGALPPECTPNIDGPNAKSVOREGSLSPFTLFCRRKPKTDCFLHP 300
DB 241 ELKERYKELTQQLPGALPPECTPNIDGPNAKSVOREGSLSPFTLFCRRKPKTDCFLHP 300
QY 299 FHATPNYTKRKNTETALNKPQCGPCYOHLGAKRFAALAAERIKTPPKRPGGRRRGRL 358
DB 299 FHATPNYTKRKNTETALNKPQCGPCYOHLGAKRFAALAAERIKTPPKRPGGRRRGRL 358
QY 301 FHATPNYTKRKNTETALNKPQCGPCYOHLGAKRFAALAAERIKTPPKRPGGRRRGRL 360
DB 301 FHATPNYTKRKNTETALNKPQCGPCYOHLGAKRFAALAAERIKTPPKRPGGRRRGRL 360

```

```

QY 359 PNNSSRPSTPTINLVESKDTDSIREAGTGTGENDKKEEKKODETSSSEANSRCQTP 418
DB 361 PNNSSRPSTPTINLVESKDTDSIREAGTGTGENDKKEEKKODETSSSEANSRCQTP 420
QY 419 KMCNTEPPENVMSCAEMSPFVLIGTYDNCALARLIGTCTCQVYEFYVESIIA 478
DB 421 KMCNTEPPENVMSCAEMSPFVLIGTYDNCALARLIGTCTCQVYEFYVESIIA 480
QY 479 PAPAEDVDTPEPRKKKKRRLMAAHCRIQLKQKDSNNHVNYPQCHPRQPCDSCPCVI 538
DB 481 PVAEDVDTPEPRKKKKRRLMAAHCRIQLKQKDSNNHVNYPQCHPRQPCDSCPCVI 540
QY 539 AONFCXPCQCSSECONRFPQGCCKAQCNTPQCPYLAVERCDPDLCTLGAADHMDSKN 598
DB 541 AONFCXPCQCSSECONRFPQGCCKAQCNTPQCPYLAVERCDPDLCTLGAADHMDSKN 600
QY 599 VSGCKNSIQRSKSHLLASDVAGWGIPIKDVQKNRPFISEYCGEIIISQDBADRGKY 658
DB 601 VSGCKNSIQRSKSHLLASDVAGWGIPIKDVQKNRPFISEYCGEIIISQDBADRGKY 660
QY 659 DKTWCSEFLFNILNDFVVDATRKGNKIRFANHNVNPNVCYAKVMVNGDHRIGIFAKRAIQT 718
DB 661 DKTCMSFLFNILNDFVVDATRKGNKIRFANHNVNPNVCYAKVMVNGDHRIGIFAKRAIQT 720
QY 719 GEBLFPDYRYSQADALKTGVIEREMEIP 746
DB 721 GEBLFPDYRYSQADALKTGVIEREMEIP 748

RESULT 11
Q4V863_XENLA PRELIMINARY; PRT; 748 AA.
ID Q4V863_XENLA
AC Q4V863;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 21-FEB-2006, entry version 6.
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22338257; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klein S.L., Straussberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.,
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22338257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Schenken C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueter K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
RA Diatchenko L., Marusita K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Tothilyuki S., Carninci F., Pirange C.,
RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heitlen E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallus D.B.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

```



"Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RA Klein S., Gerhard D.S.;  
 RC TISUE=Embryo;  
 RP NUCLEOTIDE SEQUENCE.  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 DR EMBL: BC097526; AAH97526.1; -; mRNA.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; P:DNA binding; IEA.  
 DR InterPro: IPR001005; MYD\_DNA\_bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS0280; SET; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 748 AA; 85742 MW; BD6B17AF6B7B2A6A CRC64;  
 Query Match 92.8%; Score 3758; DB 2; Length 748;  
 Best Local Similarity 92.4%; Pred. No. 1,7e-213;  
 Matches 691; Conservative 28; Mismatches 27; Indels 2; Gaps 1;  
 QY 1 MGQTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSNRQKILERTETLNDW 60  
 DB 1 MGQTGKSEKGVACRKVKSEYMRRLQKFRRADEVKSMFSNRQKIMERTETLNDW 60  
 QY 61 KORRIQPHILTSVSLAGTRCSTSDLPPTQVYPLKLTANAASVPIYMSWSPLOQNF 120  
 DB 61 KORRIQPHIMTYSLSLGRTRCFYTSIDLPFKQVYPLKLTAAVSMPIYMSWSPLOQNF 120  
 QY 121 MVEDETVLANIPLYMGDEVLDQGTFFIELIKYDQKVGDRRCGFINBEIPELVNALGQ 180  
 DB 121 MVEDETVLANIPLYMGDEVLDQGTFFIELIKYDQKVGDRRCGFINBEIPELVNALAQ 180  
 QY 181 YNDDDDDDDGDD--PEEREKQKQLEDRHDDKESRPPKFPESDKIFEAISSMFPDGTAE 238  
 DB 181 YSDYEDDDEGDDNQDDEHDITKQDDMMWKEKTLPLKFPEDKIFEAISSMFPDGTAE 240  
 QY 239 ELKEXYKELTEQQLFGALPPECTPNIIDGNNAKSVOREGSLSHFTLFCRCCKYDCLHP 298  
 DB 241 ELKEXYKELTEQQLFGALPPECTPNIIDGNNAKSVOREGSLSHFTLFCRCCKYDCLHP 300  
 QY 299 FHATPNTYKRNKTERLTALNKPQPCYOHLEGAKEFAALTAERIKTPPKRGGRRL 358  
 DB 301 FHATPNTYKRNKNEANLNDGKLCGPYCYOLLEGARFEAALTAERIKTPPKRGGRRL 360  
 QY 359 PNNSSRPSTPTINVLSEKQDSDRAGTETGGENNDKEEKKDSTSSSEANSRCQPI 418  
 DB 361 PNNSSRPSTPTINVLSEKQDSDRAGTETGGENNDKEEKKDSTSSSEANSRCQPI 420  
 QY 419 KKKENIPEPENVEWMSGAEMFVLIIGTYDNFCALIALIGTCTQRYEERFKESSIIA 478  
 DB 421 KKKPIPEPENVEWMSGAEMSLFRVLIGTYDNFCALIALISTKTCROYEERFKESSIIA 480  
 QY 479 PAAEDVUTPRKKKKRKRRLMAHCRKIOLKDDGSSNNVYVQPDHROPDSCPCVI 538  
 DB 481 PVAEDVUTPRKKKKRKRRLMAHCRKIOLKDDGSSNNVYVQPDHROPDSCPCVI 540  
 QY 539 AONFCEKQCQSSSECONRFPGCRCAQCNTKQPCYLAVRECDPLCTCGAADMDSKN 598  
 DB 541 AONFCEKQCQSSSQDNRFPGCRCAQCNTKQPCYLAVRECDPLCTCGAADMDSKN 600  
 QY 599 VSCNKGSIORGSKKILLAPSDVAGWGFIFKDPVQDNTEFISYCEBIIISODEADRRGY 658  
 DB 601 VSCNKGSIORGSKKILLAPSDVAGWGFIFKDPVQDNTEFISYCEBIIISODEADRRGY 660  
 QY 659 DKYMSPLFNLNNDPVVDATRKGNKIRFANISVNNVCACAKWVMVGDHRIQIPARAIQT 718

DB 661 DKYMSPLFNLNNDPVVDATRKGNKIRFANISVNNVCACAKWVMVGDHRIQIPARAIQT 720  
 QY 719 GEELPFDPYRSQADALKTVGIEREMEIP 746  
 DB 721 GEELPFDPYRSQADALKTVGIEREMEIP 748  
 RESULT 12  
 Q4R780 MACFA PRELIMINARY; PRT; 695 AA.  
 AC Q4R780;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 19-FEB-2006, entry version 5.  
 DE Teletis cDNA, clone: Q4a-15957, similar to human enhancer of zeste  
 DE homolog 2 (Drosophila) (EZH2), transcript variant 2.  
 OS Macaca fascicularis (Crah eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 OC Cercopithecoidea; Cercopithecinae; Macaca.  
 ON NCBI\_TaxID=9541;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15944441; DOI=10.1093/molbev/msi187;  
 RA Osada N., Hirata M., Tanuma R., Kusuda J., Hida M., Suzuki Y.,  
 RA Sugano S., Gojobori T., Shen C.-K.J., Wu C.I., Hashimoto K.;  
 RT "Substitution Rate and Structural Divergence of 5'UTR Evolution;  
 RT Comparative Analysis Between Human and Cynomolgus Monkey cDNAs.";  
 RL Mol. Biol. Evol. 22:1976-1982 (2005).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG International consortium for macaque cDNA sequencing and analysis;  
 RT "DNA sequences of macaque genes expressed in brain or testis and its  
 RT evolutionary implications.";  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 DR EMBL: AB168941; BAB01042.1; -; mRNA.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; P:DNA binding; IEA.  
 DR InterPro: IPR001005; MYD\_DNA\_bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS0280; SET; 1.  
 SQ SEQUENCE 695 AA; 79621 MW; 9DEAPB0755468660 CRC64;  
 Query Match 91.3%; Score 3698.5; DB 2; Length 695;  
 Best Local Similarity 93.0%; Pred. No. 5.1e-210;  
 Matches 694; Conservative 0; Mismatches 1; Indels 51; Gaps 2;  
 QY 1 MGQTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSNRQKILERTETLNDW 60  
 DB 1 MGQTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSNRQKILERTETLNDW 60  
 QY 61 KORRIQPHILTSVSLAGTRCSTSDLPPTQVYPLKLTANAASVPIYMSWSPLOQNF 120  
 DB 61 KORRIQPHILTSVSLAGTRCSTSDLPPTQVYPLKLTANAASVPIYMSWSPLOQNF 111  
 QY 121 MVEDETVLANIPLYMGDEVLDQGTFFIELIKYDQKVGDRRCGFINBEIPELVNALGQ 180  
 DB 121 MVEDETVLANIPLYMGDEVLDQGTFFIELIKYDQKVGDRRCGFINBEIPELVNALGQ 171  
 QY 181 YNDDDDDDDGDDPEEREKQKQLEDRHDDKESRPPKFPESDKIFEAISSMFPDGTAEEL 240  
 DB 181 YNDDDDDDDGDDPEEREKQKQLEDRHDDKESRPPKFPESDKIFEAISSMFPDGTAEEL 231  
 QY 241 KKKYKELTEQQLFGALPPECTPNIIDGNNAKSVOREGSLSHFTLFCRCCKYDCLHPFH 300

Db 232 KKKYKELTEQQLPGALPEPCTPNIDGNPAKSVQBOSLHSHFTLLFCRRCFYXDCFLHPH 291  
QY ATPNTYKRNKTETALDNKPCGPOCYOHLEGAKEFAPALTAERIKTPPKRPGRRGRILPN 360  
Db 292 ATPNTYKRNKTETALDNKPCGPOCYOHLEGAKERAPALTAERIKTPPKRPGRRGRILPN 351  
QY 361 NSSRPSTPTINVLKSDTSDREAGTETGGENNDKEBEKDETFSSSEANSRCOTPIKM 420  
Db 352 NSSRPSTPTINVLKSDTSDREAGTETGGENNDKEBEKDETFSSSEANSRCOTPIKM 411  
QY 421 KPNLEPPENWESGASAMFVLLGTYYDNFCALRIIGTCGQVYEFKXESSIIAPA 480  
Db 412 KPNLEPPENWESGASAMFVLLGTYYDNFCALRIIGTCGQVYEFKXESSIIAPA 471  
QY 481 PABVDTPPRKKKKRHLMAHCRKIOLKDGSSNNVYNYQPCDHPRQPCDSCPCVIAQ 540  
Db 472 PABVDTPPRKKKKRHLMAHCRKIOLKDGSSNNVYNYQPCDHPRQPCDSCPCVIAQ 502  
QY 541 NFCEKPCQCSSECONRPFQCKCAQCNTRYQPCYLAVERCDPLCTCGAADHMSKNVS 600  
Db 503 -----QNRFPQCRCKAQCNTRYQPCYLAVERCDPLCTCGAADHMSKNVS 549  
QY 601 CKNSIQGSKKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEITISQDEADRGKXYDK 660  
Db 550 CKNSIQGSKKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEITISQDEADRGKXYDK 609  
QY 661 YMCSEFLFNLANDFVVDATRKKNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKAIALQTGE 720  
Db 610 YMCSEFLFNLANDFVVDATRKKNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKAIALQTGE 669  
QY 721 ELFPDYRYSQADALXYGIEREMETP 746  
Db 670 ELFPDYRYSQADALXYGIEREMETP 695

RESULT 13  
Q75MP9 HUMAN PRELIMINARY; PRT; 664 AA.  
AC Q75MP9;  
DT 05-JUN-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUN-2004, sequence version 1.  
DT 21-FEB-2006, entry version 10.  
DE Hypothetical protein E2H2 (fragment).  
GN Name=E2H2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
OC Homo.  
NCBI\_TaxID=9606;  
RX NCBI [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;  
RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,  
Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walther R.,  
Wayle K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.B.,  
Fewell G.A., DeLehanty K.D., Miner T.L., Nash W.B., Cortes M., Du H.,  
Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,  
Vazirunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,  
Razzerby P., Bielicki L., Scott K., Holmes A., Harting R., Harris A.,  
Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,  
Kozlowicz-Relilly A., Leonard S., Rohlfing T., Rock S.M.,  
Rat-Lewin A.-M., Abbott A., Minx P., Maupin R., Strommat C.,  
Lafrenie P., Miller N., Johnson D., Murray J., Moesner J.P.,  
Wend M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,  
Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,  
Hickendorn M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,  
Clifton S.W., Chisoe S.L., Marra M.A., Raymond C., Haugen B.,  
Gillett W., Zhou Y., James R., Phelps K., Iadonito S., Bubb K.,  
Simms B., Levy R., Glendinning J., Kaul R., Kent W.J., Furey T.S.,  
Bartesch R.A., Brent M.R., Keibler E., Flicker P., Bork P., Suyama M.,  
Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,  
Rddy S.R., McPherson J.D., Olson M.V., Eichler E.R., Green E.D.,  
Waterston R.H., Wilson R.K.;

RT "The DNA sequence of human chromosome 7.,"  
RL Nature 424:157-164(2003).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Bauer C., Holmes A., Mead K.;  
RT "The sequence of Homo sapiens PAC clone Rps-1151W5.,"  
RN Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RP [3]  
RP NUCLEOTIDE SEQUENCE.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RA Wilson R.;  
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivative license  
CC -----  
CC EMBL; AC006323; AA502035.1; -; Genomic DNA.  
DR Ensembl; ENSG00000106462; Homo sapiens.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR001005; Myb\_DNA\_Bd.  
DR InterPro; IPR01214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00717; SANT; 2.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS02080; SET; 1.  
KW Hypothetical protein.  
FT NON TER 1  
SQ SEQUENCE 664 AA; 75478 MW; 585FCE1B5D8651C8 CRC64;  
Query Match 89.6%; Score 3631; DB 2; Length 664;  
Best Local Similarity 100.0%; Pred. No. 4.7e-206;  
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 CSVTSDDLPPTQVPLPLTLNAVASVPTMYSKSPLOQNFWEDEVYLNIPMGREVDOD 142  
Db 1 CSVTSDDLPPTQVPLPLTLNAVASVPTMYSKSPLOQNFWEDEVYLNIPMGREVDOD 60  
QY 143 GTFIEELIKYVDGVDHREGFINDELFVELVNAQYNDDDDDGDDDEBEKOKD 202  
Db 61 GTFIEELIKYVDGVDHREGFINDELFVELVNAQYNDDDDDGDDDEBEKOKD 120  
QY 203 LEDHRDDKESRPPKPSDKIFEAISMFDPKGTAEELKEKYKELTEQQLPGALPPECTP 262  
Db 121 LEDHRDDKESRPPKPSDKIFEAISMFDPKGTAEELKEKYKELTEQQLPGALPPECTP 180  
QY 263 NIDGPNAKSVQBOSLHSHFTLLFCRCGRKIDCFLHPHATPNYTKRNKTETALDNKPCG 322  
Db 181 NIDGPNAKSVQBOSLHSHFTLLFCRCGRKIDCFLHPHATPNYTKRNKTETALDNKPCG 240  
QY 323 QCYOHLGAKERAPALTAERIKTPPKRPGRRGRLLPNNSRPSTPTINVLKSDTSDR 382  
Db 241 QCYOHLGAKERAPALTAERIKTPPKRPGRRGRLLPNNSRPSTPTINVLKSDTSDR 300  
QY 383 EAGTETGENDNKEBEKDETFSSSEANSRCOTPIKKNLEPPENWESGASAMFV 442  
Db 301 EAGTETGENDNKEBEKDETFSSSEANSRCOTPIKKNLEPPENWESGASAMFV 360  
QY 443 LIGTYNFCALRIIGTCGQVYEFKXESSIIAPAPADVDTPPKKKRKLMAAH 502  
Db 361 LIGTYNFCALRIIGTCGQVYEFKXESSIIAPAPADVDTPPKKKRKLMAAH 420  
QY 503 CRKIOLKDGSSNNVYNYQPCDHPRQPCDSCPCVIAQNFCEKPCQCSSECONRPFQCR 562  
Db 421 CRKIOLKDGSSNNVYNYQPCDHPRQPCDSCPCVIAQNFCEKPCQCSSECONRPFQCR 480  
QY 563 KAQCNTRYQPCYLAVERCDPLCTCGAADHMSKNVSKNCSIQGSKKHLLAPSDVA 622  
Db 481 KAQCNTRYQPCYLAVERCDPLCTCGAADHMSKNVSKNCSIQGSKKHLLAPSDVA 540

QY 623 GNGIFKDPVQKNEPISBYCGELISQDEADRGKVDYDKMCSFLFNLIANDPVVADTRKGN 682  
 DB 541 GNGIFKDPVQKNEPISBYCGELISQDEADRGKVDYDKMCSFLFNLIANDPVVADTRKGN 600  
 QY 683 KIRPANHSVNPNCYAKVMVNGDHRIGIFAKKAIQTGEELFPDYRSQADALKYGIERE 742  
 DB 601 KIRPANHSVNPNCYAKVMVNGDHRIGIFAKKAIQTGEELFPDYRSQADALKYGIERE 660  
 QY 743 MEIP 746  
 DB 661 MEIP 664  
 RESULT 14  
 Q75W00\_HUMAN PRELIMINARY; PRT; 669 AA.  
 AC Q75W00;  
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 05-JUL-2004, sequence version 1.  
 DT 21-FEB-2006, entry version 10.  
 DE Hypothetical protein EZH2 (Fragment).  
 GN Name=EZH2;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE:22737999; PubMed:12853948; DOI=10.1038/nature01782;  
 RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,  
 Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,  
 Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,  
 Fewell G.A., Delahunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,  
 Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,  
 Vazirunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,  
 Zaretsky P., Bielicki L., Scott K., Holmes A., Hartings R., Harris A.,  
 Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,  
 Kozlowski-Reilly A., Leonard A., Mink P., Maupin R., Stromwater C.,  
 Tin-Wollam A.-M., Abbott A., Mink P., Maupin R., Stromwater C.,  
 Latreille P., Miller N., Johnson D., Murray J., Moesner J.P.,  
 Mendl M.C., Yang S.-P., Schultze B.R., Wallis J.W., Spieth J.,  
 Bieri T.A., Nelson J.O., Bertozzi N., Wohlmann P.E., Cook L.L.,  
 Hickenbotham M.T., Bladed J., Williams D., Bedell J.A., Mardis E.R.,  
 Clifton S.W., Chiswick S.L., Marra M.A., Raymond C., Haugen B.,  
 Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubb K.,  
 Simms B., Levy R., Glendinning J., Kaul R., Kent W.J., Frey T.S.,  
 Baerbach R.A., Brent M.R., Keibler B., Flicker P., Bork P., Suyama M.,  
 Bailey J.A., Portnoy M.E., Torrents D., Chitwalia A.T., Gish W.R.,  
 Bddy S.R., McPherson J.D., Olson M.V., Richter E.E., Green E.D.,  
 Waterston R.H., Wilson R.K.;  
 RT "The DNA sequence of human chromosome 7.";  
 RL Nature 424:157-164(2003).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Bauer C., Holmes A., Mead K.;  
 RT "The sequence of Homo sapiens PAC clone RP5-1151M5.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Waterston R.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Wilson R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial license  
 CC EMBL: AC006323; AAS02036.1; -; Genomic DNA.  
 DR Ensembl: ENSG00000106462; Homo sapiens.  
 GO: GO:0005634; C:nucleus, IEA.

DR GO: GO:0003677; F:DNA binding, IEA.  
 DR InterPro: IPR001005; Mtd\_DNA\_bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS00280; SET; 1.  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 669 AA; 76132 MW; 413165183ABEF7D1 CRC64;  
 Query Match 89.1%; Score 3610.5; DB 2; Length 669;  
 Best Local Similarity 99.1%; Pred. No. 7.6e-205;  
 Matches 663; Conservative 0; Mismatches 1; Indels 5; Gaps 1;  
 QY 83 CSVTSDDLPPQVPIPLKLTNAVASVPIVMSPLQONFVDETVLNI PYMGDEVLDQD 142  
 DB 1 CSVTSDDLPPQVPIPLKLTNAVASVPIVMSPLQONFVDETVLNI PYMGDEVLDQD 60  
 QY 143 GTFIELIKNTDGVHGRGEGFINDEIFVELVNAAGQYNDDDDDDDDDDEBEREKXD 202  
 DB 61 GTFIELIKNTDGVHGRGEGFINDEIFVELVNAAGQYNDDDDDDDDDDEBEREKXD 120  
 QY 203 LEDHRDDKSRPPKPFSDKIFBAISSWFPDKTAELKEKYLTEBOQLPGALPPECCTP 262  
 DB 121 LEDHRDDKSRPPKPFSDKIFBAISSWFPDKTAELKEKYLTEBOQLPGALPPECCTP 180  
 QY 263 NIDGPNKSVOREGSLSHFHTLFCRCPCFYDCFLH-----PFHATPNTYKRNKTETALDN 317  
 DB 181 NIDGPNKSVOREGSLSHFHTLFCRCPCFYDCFLHRCNKNVCFHATPNTYKRNKTETALDN 240  
 QY 318 KPCGPOCYQHLEGAKEFPAALTAIRIKTPPRPGRRGRGLPNNSSRSTPTIIVLESKD 377  
 DB 241 KPCGPOCYQHLEGAKEFPAALTAIRIKTPPRPGRRGRGLPNNSSRSTPTIIVLESKD 300  
 QY 378 TDSRREGTEGNNNDGSEBEKDETSSEANSRSCOTPIKMKPNIEPPNVEVMSGABA 437  
 DB 301 TDSRREGTEGNNNDGSEBEKDETSSEANSRSCOTPIKMKPNIEPPNVEVMSGABA 360  
 QY 438 SMFVLTIGTYVDFNCALIRLIGTTCQVYEFVYESSIIAPAPAEVDVTPPRKKGRGR 497  
 DB 361 SMFVLTIGTYVDFNCALIRLIGTTCQVYEFVYESSIIAPAPAEVDVTPPRKKGRGR 420  
 QY 498 LMAAHCRTQLKXGSSNHTVNYOPCDHPPOPCSSCPCVIAONFCERCCSSCCQNR 557  
 DB 421 LMAAHCRTQLKXGSSNHTVNYOPCDHPPOPCSSCPCVIAONFCERCCSSCCQNR 480  
 QY 558 PGCRCKAQCNKQPCVIAVRECDPDLCTGGAADHMSKRVSCNCSIQGSKKHLLA 617  
 DB 481 PGCRCKAQCNKQPCVIAVRECDPDLCTGGAADHMSKRVSCNCSIQGSKKHLLA 540  
 QY 618 PSDVAGNGIFPKDPVQKNEPISBYCGELISQDEADRGKVDYDKMCSFLFNLIANDPVVDA 677  
 DB 541 PSDVAGNGIFPKDPVQKNEPISBYCGELISQDEADRGKVDYDKMCSFLFNLIANDPVVDA 600  
 QY 678 TRKNKTRPANHSVNPNCYAKVMVNGDHRIGIFAKKAIQTGEELFPDYRSQADALKYV 737  
 DB 601 TRKNKTRPANHSVNPNCYAKVMVNGDHRIGIFAKKAIQTGEELFPDYRSQADALKYV 660  
 QY 738 GIEREMIP 746  
 DB 661 GIEREMIP 669  
 RESULT 15  
 Q90WP4\_TETNG PRELIMINARY; PRT; 759 AA.  
 AC Q90WP4;  
 DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-DEC-2001, sequence version 1.  
 DT 21-FEB-2006, entry version 20.  
 DE EZH2 homolog.  
 GN Name=ezh2;

OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 NCBI\_TaxID=99863;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Blood;  
 RA MEDLINE=22269943; PubMed=1368471; DOI=10.1073/pnas.202284199;  
 RA Dasilva C., Hadji H., Ozouf-Costaz C., Nicand S., Jalllon O.,  
 RA Wellesbach U., Roest Crolius H.;  
 RT "Remarkable compartmentalization of transposable elements and  
 RT pseudogenes in the heterochromatin of the Tetraodon nigroviridis  
 RT genome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13636-13641(2002).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC -----  
 DR EMBL: AJ131481; CAC86146.1; -; Genomic\_DNA.  
 DR HSSP: Q8X225; 1ML9.  
 DR Ensembl: GSTENG0007024001; Tetraodon nigroviridis.  
 DR GO: GO:0005634; C:nucleus, IEA.  
 DR GO: GO:0003677; F:DNA binding, IEA.  
 DR InterPro: IPR01005; MYD\_DNA\_Bd.  
 DR InterPro: IPR01214; SET.  
 DR Pfam: PF00856; SET, 1.  
 DR SMART: SM00717; SANT, 2.  
 DR SMART: SM00317; SET, 1.  
 DR PROSITE: PS50280; SET, 1.  
 SQ SEQUENCE 759 AA; 86834 MW; DBA5DE76922P640 CRC64;  
 Query Match 83.8%; Score 3394.5; DB 2; Length 759;  
 Best Local Similarity 83.2%; Pred. No. 5e-192;  
 Matches 632; Conservative 53; Mismatches 52; Indels 23; Gaps 10;  
 4 TGKSEKGPVCKWRKVKSEYMLRQLKPRADAEVYKSPSSNRKILERTILNQEWKOR 63  
 4 TGKSEKGPACMKRKRKSEYMLRQLKPRADAEVYKSPSSNRKILERTILNQEWKTR 63  
 64 RIQPHILTSVSLRGTECVTSDL-DEPTQVPLKTLNVAASVPIVMSPLQONFMV 122  
 64 RIQPHIMTSVSLRGTECVTSDFSSEFPQVPLKTLNVAASVPIVMSPLQONFMV 123  
 123 EDETVAHNI PYMGDEVLDDGTFLBELIKNYDGKVGDRGCGFINDEIFVELVNAAGQYN 182  
 124 EDETVAHNI PYMGDEVLDDGTFLBELIKNYDGKVGDRGCGFINDEIFVELVNAAGYS 183  
 183 DDDDDGDDGDPBEREEK-----OKLDEHRDD---KESRP---PKPPSDKIFEAIS 228  
 184 DNEDEEEHEHDPKVDKIDLCDSKGPRDGKGLVNNGRSGADSKKPPSDKIFEAIS 243  
 229 SMPDPKGAEBELKEXKELTEQQLPGALPPECTPNIIDGPNAKSVORBSLSFHTLFCRR 286  
 244 AMPDPKGSIEBELKEXKELTEQOMGALPPECTPNIIDGPNAKSVORBSLSFHTLFCRR 303  
 289 CFKYDCFLHPHATPNTYKRNKTELTALDNKPCGPOCYOL--EG-AKEFAALTEARIKT 345  
 304 CFKYDCFLHPHATPNTYKRNKLENLVDSKPCGLDVMYLVODGMASEYAAAG--AERTKT 361  
 346 PPKRPGRRGRRLPNNSRSPSTPTINVLSEKDTSDREAGTETGGENNDEBEKEDETS 405  
 362 PSKRPAGRRGRQPPSSSRPSTPTVS--SESKDADSDRE--GSKEDERENKEDDKDENT 419  
 406 SSSEANSCTPIKKKPIEPPEVWEGASWFRVLTGYNDPFCALRLIGTKTCRQ 465  
 420 SSSSEANSCTPIPKK--LVGEAEAVDMMGAESLFRVLIGTYNDFCALRLIGTKTCRQ 478  
 466 VYEFVKESIIAPAEADVTPPKKKRKRRLMAHCRKIQLKKDSSNHVYNYQPCDH 525  
 479 VYEFVKESIIAPAEADVTPPKKKRKRRLMAHCRKIQLKKDSSNHVYNYQPCDH 538

QY 526 PRQPCDSSCPVIAQNTCEKFCQCCSSSECQNRFPQCRCKAQONTQCCCYLAARECDPDL 585  
 DB 539 PRQPCDSSCPVIAQNTCEKFCQCCSSSECQNRFPQCRCKAQONTQCCCYLAARECDPDL 598  
 QY 586 LTGGAADHMSKXVSCNKGSIORGSKGHLAPSDVAGWGIPIKD PVQXNFEISEYCGEI 645  
 DB 599 LTGGAADHMSKXVSCNKGSIORGAKGHLAPSDVAGWGIPIKEPVQXNDFISEYCGEI 658  
 QY 646 ISQDEADRGKVDKXNCSFLFNLNDPVVDATRKNGKIRIFANHSVNPNCYAKVMVNGD 705  
 DB 659 ISQDEADRGKVDKXNCSFLFNLNDPVVDATRKNGKIRIFANHSVNPNCYAKVMVNGD 718  
 QY 706 HRIGIFAKRAIQTSEELFPDTRYSQADALKYVGIRERMEI 745  
 DB 719 HRIGIFAKRAIQTSEELFPDTRYSQADARXYGIEREMEI 758

Search completed: August 11, 2006, 22:28:02  
 Job time : 207.794 secs



Db 241 KEKKEKELTEQOLPGALPBECPPTNIDGNPAKSVQREQSHSFHTLFCRCRCFYDCFLHPFH 300  
Qy 301 ATPNTYRKKTETALDNKPCGPOCYOHLBGAKEFAALTAERIKTPPRKPGRRGRGLPN 360  
Db 301 ATPNTYRKKTETALDNKPCGPOCYOHLBGAKEFAALTAERIKTPPRKPGRRGRGLPN 360  
Qy 361 NSSRPSTPTINVLBSKOTSDREAGTETGGENNDKEBEKKOETSSSEANSRCQTP1KM 420  
Db 361 NSSRPSTPTINVLBSKOTSDREAGTETGGENNDKEBEKKOETSSSEANSRCQTP1KM 420  
Qy 421 KPNIPEPENWMSGAEASMERVLIGTYYNDFCALIARLIGTKCRQYVEFRVKESSIIAPA 480  
Db 421 KPNIPEPENWMSGAEASMERVLIGTYYNDFCALIARLIGTKCRQYVEFRVKESSIIAPA 480  
Qy 481 PAEDVTPPRKKRKRHLMAHCRKIQLKDGSSNHYNYNQPCHPRQPCDSSCPVIAQ 540  
Db 481 PAEDVTPPRKKRKRHLMAHCRKIQLKDGSSNHYNYNQPCHPRQPCDSSCPVIAQ 540  
Qy 541 NFCEKFCQCSSSECONRPFPGCRCAQCNTKQCPCTLAVERCDPDLCTCGAADHMSKNVS 600  
Db 541 NFCEKFCQCSSSECONRPFPGCRCAQCNTKQCPCTLAVERCDPDLCTCGAADHMSKNVS 600  
Qy 601 CKNCSIORGSKKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRRGKYDK 660  
Db 601 CKNCSIORGSKKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRRGKYDK 660  
Qy 661 YMCSEFLFNANDFYVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRALQTGE 720  
Db 661 YMCSEFLFNANDFYVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRALQTGE 720  
Qy 721 ELFPDYRYSQADALKYVIGIEREMEIP 746  
Db 721 ELFPDYRYSQADALKYVIGIEREMEIP 746  
RESULT 2  
US-10-104-047-3162  
Sequence 3162, Application US/10104047  
Patent No. 6943241  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. 6943241el full length cDNA  
FILE REFERENCE: H1-A0105  
CURRENT APPLICATION NUMBER: US/10/104, 047  
PRIOR FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3162  
LENGTH: 707  
TYPE: PR1  
ORGANISM: Homo sapiens  
US-10-104-047-3162  
Query Match 94.1%; Score 3812.5; DB 2; Length 707;  
Best Local Similarity 94.6%; Pred. No. 0;  
Matches 706; Conservative 0; Mismatches 1; Indels 39; Gaps 1;  
Qy 1 MGQGGKSEKGPVCKRKVSEYWRRLQKFRRADEVKSMFSSNRQKILERTETLNDQEW 60  
Db 1 MGQGGKSEKGPVCKRKVSEYWRRLQKFRRADEVKSMFSSNRQKILERTETLNDQEW 60  
Qy 61 KORRIOPVHILTSVSLAGTRECSTSDLPPTQVITPLKTMANAVAVPIWWSPLQQNF 120  
Db 61 KORRIOPVHILTSVSLAGTRECSTSDLPPTQVITPLKTMANAVAVPIWWSPLQQNF 120  
Qy 121 MVEDEVVLANIPYMGDEVLDQDGTFIIEBLINNYDGKVNDRGCGFINDEIFVELVNALGQ 180  
Db 83 -VEDEVVLANIPYMGDEVLDQDGTFIIEBLINNYDGKVNDRGCGFINDEIFVELVNALGQ 141  
Qy 181 YNDDDDDDDGDDPEEREREKQDLBDHRDKESRPPRKFPDRIEALISMFPDQGTABEL 240

Db 142 YNDDDDDDDGDDPEEREREKQDLBDHRDKESRPPRKFPDRIEALISMFPDQGTABEL 201  
Qy 241 KEKKEKELTEQOLPGALPBECPPTNIDGNPAKSVQREQSHSFHTLFCRCRCFYDCFLHPFH 300  
Db 202 KEKKEKELTEQOLPGALPBECPPTNIDGNPAKSVQREQSHSFHTLFCRCRCFYDCFLHPFH 261  
Qy 301 ATPNTYRKKTETALDNKPCGPOCYOHLBGAKEFAALTAERIKTPPRKPGRRGRGLPN 360  
Db 262 ATPNTYRKKTETALDNKPCGPOCYOHLBGAKEFAALTAERIKTPPRKPGRRGRGLPN 321  
Qy 361 NSSRPSTPTINVLBSKOTSDREAGTETGGENNDKEBEKKOETSSSEANSRCQTP1KM 420  
Db 322 NSSRPSTPTINVLBSKOTSDREAGTETGGENNDKEBEKKOETSSSEANSRCQTP1KM 381  
Qy 421 KPNIPEPENWMSGAEASMERVLIGTYYNDFCALIARLIGTKCRQYVEFRVKESSIIAPA 480  
Db 382 KPNIPEPENWMSGAEASMERVLIGTYYNDFCALIARLIGTKCRQYVEFRVKESSIIAPA 441  
Qy 481 PAEDVTPPRKKRKRHLMAHCRKIQLKDGSSNHYNYNQPCHPRQPCDSSCPVIAQ 540  
Db 442 PAEDVTPPRKKRKRHLMAHCRKIQLKDGSSNHYNYNQPCHPRQPCDSSCPVIAQ 501  
Qy 541 NFCEKFCQCSSSECONRPFPGCRCAQCNTKQCPCTLAVERCDPDLCTCGAADHMSKNVS 600  
Db 502 NFCEKFCQCSSSECONRPFPGCRCAQCNTKQCPCTLAVERCDPDLCTCGAADHMSKNVS 561  
Qy 601 CKNCSIORGSKKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRRGKYDK 660  
Db 562 CKNCSIORGSKKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRRGKYDK 621  
Qy 661 YMCSEFLFNANDFYVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRALQTGE 720  
Db 622 YMCSEFLFNANDFYVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRALQTGE 681  
Qy 721 ELFPDYRYSQADALKYVIGIEREMEIP 746  
Db 682 ELFPDYRYSQADALKYVIGIEREMEIP 707  
RESULT 3  
US-09-589-892B-11  
Sequence 11, Application US/09589892B  
Patent No. 6689583  
GENERAL INFORMATION:  
APPLICANT: Jenuwein, Thomas  
APPLICANT: Laible, Gotz  
APPLICANT: O'Carroll, Donal  
APPLICANT: Eisenhaber, Frank  
APPLICANT: Rea, Stephen  
TITLE OF INVENTION: Chromatin-Regulator Genes  
FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/09/589, 892B  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945, 988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
PRIOR FILING DATE: 1995-05-10  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 11  
LENGTH: 760  
TYPE: PR1  
ORGANISM: Drosophila melanogaster  
US-09-589-892B-11  
Query Match 52.6%; Score 2131.5; DB 2; Length 760;  
Best Local Similarity 53.9%; Pred. No. 4; 1e-177;  
Matches 433; Conservative 87; Mismatches 171; Indels 113; Gaps 17;  
Qy 8 SEKGPVCKRKVSEYWRRLQKFRRADEVKSMFSSNRQKILERTETL---NQEWQR 64

Db	3	STKPPKMKRKYVSEYIKIRIQGRYKRADEIKEMIRMDENHNHVODLUCESKYMOKAP	62
Qy	65	IQPVHILTVSSILGTRRECVTSDDLPEPT--QVTPKTLMAVASVPIMYSPLQONFMV	122
Db	63	YDPHV-----DCVGRAEVTSYNGIPSGQPKVPCIBINAVTPIPTMYTMAPTQONFMV	116
Qy	123	EDETVLNIPYMGDEVLDODCTFIIEELIKNYGKNGHDECCFINDELFUELYAL-----	178
Db	116	EDETVLNIPYMGDEVLDKDKGFIIEELIKNYGKNGHDXDPSEFMDALFVELYHLMKRSY	178
Qy	179	-----GQYND-----DDDDGGDPB--BREKQXOLE-----	205
Db	176	SKELEBAAPSTSTAIKTEPLAKSKQGBDGVVDVADCSFMKLEKTSKGLDIVEKGE	233
Qy	206	-----HRDKESRPPRK-----PSSDKIFEAISMFDPKQTAEBLKYKELTEOQLP	253
Db	236	TEBVEATEDADVPAVEVDKLPFPAPAIIFQIISANFPDKQTAQELKELYTELTHEQDP	299
Qy	254	GALPPECTPINIDGPNKASVOREGSLASTFTLPERCFKTCIDCLHP--HAIPNTYKAKT	311
Db	296	-BRPQCTPINIDGPKAESVREKRTWASFTLLPRRCFKTDCFLHLOGHAGPNIQKRYP	353
Qy	312	ETALDNPCPGPOCYHLEGAKEPAAALTBERIKTPKRGKRRGRRLPNNSRSTPTIN	373
Db	355	ELKFPAPKCSNSCTMLIDGKKEKLADS---KTP-----	386
Qy	372	VLESKDYSDREAGTEGTGGENNDK-----BEEBKQETSSEANSRCOTPIKMKPI	424
Db	387	-----IDSCNEASBSBDSNDSNQPFENQPMHNENSDNGLTNSAAVABINSIMAGMNI	440
Qy	425	EPENNVWSGAEASMFERYLIGTYINDPFCALARIIGTKCRQYERFVRESSTIAPAPAD	484
Db	441	TSIQCV--WTGADALYKVLHKVYLKNYCAIAHNMULTKCRQYERFAQKED--AESFSD	496
Qy	485	VD---TEPRKKKKRHLMAHCKRIOLKKDSSNHYVYQPCDHROGQDSSCPVIAON	543
Db	497	LRODFTPRKKKKKQORWMSLHCKRIOLKKDSSNHYVYTRCDHGHGPCDMNCSCTIOTN	556
Qy	542	FCEKFCQSSSECONRFPGCKCKAQCNWTKOCPCLAVRECDPLCTCGAADHMSKNVSC	601
Db	557	FCEKFCQSSSDCONRFPGCKCKAQCNWTKOCPCLAVRECDPLCOACG-ADQFKLTKTIC	615
Qy	602	KNCISIQGSKKHLILAPSDVAGMGIFIKDPVQKNEFISYCGEIIISQDEADRGKRYDY	661
Db	616	KNVCVQGLKHLKHLMAPSDIAGWIGIFLKEGAQKNEFISYCGEIIISQDEADRGKRYDY	675
Qy	662	MCSLEPLANDPVVDAATRKGNKIRFANHSVNPCCAKYMMNNGDRIGIFAKRAIOPGEE	722
Db	676	MCSLEPLANDPVVDAATRKGNKIRFANHSINPCYAKYMMVYGDHRIGIFAKRAIOPGEE	735
Qy	722	LFPDYRSQADALKYVGIEREMI 745	
Db	736	LFPDYRIGPTEOLKVGIEREMI 759	
RESULT 4			
US-09-699-266A-13			
Sequence 13. Application US/09699266A			
Patent No. 6559354			
GENERAL INFORMATION:			
APPLICANT: Ma, Hongchang			
APPLICANT: Morakinyo, Layo O.			
APPLICANT: Odell, Joan T.			
APPLICANT: Orozco Jr., Emil M.			
APPLICANT: Rafaleki, J. Antoni			
TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS			
FILE REFERENCE: BB1164 US NA			
CURRENT APPLICATION NUMBER: US/09/699,266A			
CURRENT FILING DATE: 2000-10-27			
PRIOR APPLICATION NUMBER: PCT/US99/08385			
PRIOR FILING DATE: 1999-04-16			
PRIOR APPLICATION NUMBER: 60/083,212			

```

      PRIOR FILING DATE: 1998-04-27
      NUMBER OF SEQ ID NOS: 13
      SOFTWARE: Microsoft Office 97
      SEQ ID NO 13
      LENGTH: 856
      TYPE: PRT
      ORGANISM: Arabidopsis thaliana
      US-09-699-266A-13

Query Match      20.5%; Score 828.5; DB 2; Length 856;
Beet Local Similarity 28.8%; Pred. No. 2,8e-63;
Matches 251; Conservative 113; Mismatches 267; Indels 241; Gaps 368

Oy      25 RLQLKRRFRADVSKM---FSSNOKILERILLNQGMQRRLQPHILLTSVSLAGTR 81
Db      35 RLSELKRKIKQGEVRSIKKKFFANNKVV-----DAHVSP--FSSAASRATA 79
Oy      82 EGSVTSDDLDPPTVILPLKTLNAAVASPIVWSPLQGNFVDETVLNI-----P 132
Db      80 EDNGNSNMLSSRRKMLCTKNG-----FSHGVDRODYVPTKQVLSASVGLPIAERIIP 133
Oy      133 YMGDEVLDQDGTPIEELIKNYDKYHGRCGFINDKI FVEILNALGQYND--DDDG 190
Db      133 YTTWIFLDRNQRYAA-----DQSVGRR-----QIYYE-----QHGGETLIGDSE 172
Oy      191 DDPBEREKOKLIEDHRDKESSRPPRKIPPSDKIFPAISMSFPDKTAEELK----- 245
Db      174 EEPEPEBEKKKEFESEG-----EDSIILWLIQGEY--GMGEFVODALCOLLSV 216
Oy      242 -----EKYVELTEQQLPALPPECTPNIDGNAK-SVQRO-----SLSHPTLFCRCF 290
Db      217 DASDILIRNYELKQD-----KQNTBEFSNSGFFLGISLKGIGALDSDFNLPCKRCL 270
Oy      291 KYDCFLHPHFAFNTYKRN--TETALDNKECPQCYOHLGAKS-----FAA----- 336
Db      271 VFPCRLLHGCSQPLISASEKQPYMSDYBGRKPKCSHGCTQLKAVAREVETCSNFSABE 330
Oy      337 -----ALTAER---ITPPKRPQ-----GRRR----- 355
Db      331 KASEBECRAVSADYVPHAAAGSVLSQVEKTDIGIKNVDSGVEBENGIRKREVPILKD 390
Oy      356 -GLPNNSRRP-----STPIVNLSEKTDSDRBAEGTENGENNKEBEK 400
Db      391 SNDLPMLNSKKQKTAASDTKMSFVNSVPSLD--QALDSTKGDQGTIDNKVNRSEADAK 446
Oy      401 K-----DETSSSEANSRCQTP-----IKKKPIREPPENVWESGAESMFRVLIG 445
Db      449 EYGEPIPTDMSYHDGSSICQPHHSGNGAIIIAEISEISR--STENNPILKDIYLGVE 506
Oy      446 TTYDNFCAIAR--LIGTKTCQOYEF-RVKESSIIAPA-----PAEDVD--TP 488
Db      507 IFRNSCLTARNLISGLTKCLDVSNVMEENEVSFRRSSTPNLLDGRGTPGANDNEVP 566
Oy      489 PR---KKKKRRLW-----AAHCKIQLKDDGSNNVNYVYOCBHPRCQDSSCCPVI 538
Db      567 PKRLRLRRRKGRKRLKYSTKSAGHSVWKRLAGGNQGSCKQYTPCG-CLSMGCKGCPLT 625
Oy      539 AONFCEKFCOCSBCCNRPFGRC--KAOCNTKQCPCYLAVBECPDCLTGCADHMDSK 597
Db      626 NEFCCEKYGCCSASCNNRFRGCHCAKSOCSRQCCPFAAGBECPDVCRNC-----W--- 677
Oy      598 NVSCNKSCTI---ORG-----SKHLLAPSDVAGWGIFIKDPVQKNFISBY 641
Db      678 -VSCGDSGISGEAPRBGEQCGNMRLLLRQORILLGKSDVAGWGAFILNSVSKREYIGEY 746
Oy      642 CGEITIQDADRDRGKTYDKMCSFLFANLNDPVYDATRKGNKIRFANHSVNPNCYAYVM 700
Db      737 TGEILISHHREDRGRKITYRPAANSFFLPDLNDYVILDAQRKDKLKFANHSAKPNCYAKVF 766
Oy      702 VNGDHRIGIFAKKAIOTGEELFPDYRYGQADA 733
Db      797 VAGDHRVGIIFANERIEASEHELFTYDRIYGPDA 828

```

```

RESULT 5
US-09-699-266A-12
Sequence 12, Application US/09699266A
Patent No. 6559354
GENERAL INFORMATION:
APPLICANT: Ma, Hongchang
APPLICANT: Morakinyo, Layo O.
APPLICANT: Odell, Joan T.
APPLICANT: Orozco Jr., Emil M.
APPLICANT: Rafalecki, J. Antoni
TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS
FILE REFERENCE: BB1164 US NA
CURRENT APPLICATION NUMBER: US/09/699,266A
CURRENT FILING DATE: 2000-10-27
PRIOR APPLICATION NUMBER: PCT/US99/08395
PRIOR FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/083,212
PRIOR FILING DATE: 1998-04-27
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 902
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-699-266A-12

```

Query Match	19.6%	Score 795.5	DB 2	Length 502;
Best Local Similarity	27.0%	Pred. No. 2.3e-60;		
Matches 251;	Conservative 110;	Mismatches 274;	Indels 295;	Gaps 35

Qy	9	KKGJPCWKRKRKXSEYWMRLRQJLFRPARADEVSMFSNSOKLIERFEILNOEWKQRIQV	68
Db	24	ERGPA--SKSEYSEVIESLKKGLADRCISIKRIDENKGNLPATIQSTFR	71
Qy	69	HILTSVSSLRGTRECVTSU-----DPTQVILPKTLN-----AVASV	107
Db	72	-----SMERKGSCKDSGSDLLVYKRQRDPSGKSGSIDESNNNRVYEDBPASSGMYQSGSV	125
Qy	108	PIMYWSPLQONFWMEDETVLNIIPYMDDEVLDQOCTPIBELIKNYDKVHSDRECGPIN	167
Db	126	FVKISLRPIKM-----PDIKLSFYTTWVFLDRNQRMTE-----DOSVGR-----	167
Qy	168	DEIFVELYNALGOYNDDDDDDDGDDPPEEREKOKLBDHRDDKESRPPRKF-----	218
Db	168	-RIYDQNGEALICSDSEBBADIDDEEK-----RDPLF--PEDYIIRMTLEQL	213
Qy	219	-PSDKIFPAISMFPDKGTAELEKRYKELTEQOLGALPPTCTYNIDGNPAKSV--QR	274
Db	214	GLSDSVLAELMS-FLSRST--RIKARHGLVMKEK-----EVSSEGDQOASSLNTNDM	264
Qy	275	EOSLSHFLLFCRRCFCYKDCFLH-----PFIATENTYRKROTETALDNKPCGP	322
Db	265	EGALDSFONLFCRRCLVDFPCRHLGSGODLIPPAKRPAPWC-----PYDENILTCGA	315
Qy	323	QCYQHLBEAKERPAALTYAE-----RITPPKRRPGRRRGRLPNN--SSRSTPTI	370
Db	316	NCYKTLKSGRPGYGTIEGKTGTSSDGAQITTYTKRSSKLGKRPKTPFESASASBEK	375
Qy	371	NVLBS-----KDTUSD-----REACTETGGENN-----	393
Db	376	CALERTSDSENGLQDUTNSDDKYSSPKVYSGRRVERKKNKNVARVRPRTQKQKTEA	435
Qy	394	-----DKEEBKODETSSS-----SEANSQCTPIPKKPV-----	423
Db	436	SDSDSIASGSCSPSPDAKXMDNEDATSSQKHVYSGNSGKSRNGGYPARVSNNSVVDQVY	495
Qy	424	-----IEPPENVV-----WSGAEASMEFVLIGTYDNC	452
Db	496	CQSNVEVASELDAPGDSDELKKEEPMGTYSRRLATNKLMLREKSLDFPKXGVEIRGMNSC	555
Qy	453	AIAR--LIGTYTCROVYEFRV-----KSSITAPAPAEVDYTPPRKKRK	495

```

Db      556  LIARNLISGFKSCMEVFQYMTSEKKA5FPGDGLNPSSKSPFINGMVMNNQVRRSRF  615
Qy      496  HR-----LW--AAHC--KICLAKDQSSNHYNYNPPCHPRQ-----PCDS  532
Db      616  LRRGRVRLKYTWMSAAHYSIRKRIITKKQ-----OPC--HOFNPNCKIAAGK  663
Qy      533  SCPCVIAONFCEKFCQCSSECONREPGRC--KAQNTKQCPYLAVECDPDLCTGAA  591
Db      664  EPCCLLNGTCCEKYCGCPKSCNRRPGCHCAKSCQSRHQCPFAADECCPDVVCNC--  720
Qy      592  DHK-----DSKNSCKKCSQIQRASKKHLLAPSDVAGWGIPTKDPQKNEFI  638
Db      721  --VWIGDGSGLVGPSSRGDNIECRNMKLLAKQQRVLVLGISDVSGWGAFLKNSYSKREYL  778
Qy      639  SEYCGEIIISQDEADRRGRKYVCKYMC5FLFNINLNDPVYDATRKXKILIRPANH5VNPNCYAK  698
Db      779  GEYTGELIISHKADLRKKGKLYDRENSFLFNINDPVLVDAYRKGDKLKFPAHNSPEPNCYAK  838
Qy      699  VMMVNGDHRIGIFAKRAIQTGEBELPPDRY  728
Db      839  VIMVAGDHRVGIFAKERILLAGBELYYDRY  868

```

RESULT 6  
US-09-699-266A-9  
! Sequence 9, Application US/09699266A

```

; GENERAL INFORMATION:
; APPLICANT: Ma, Hongchang
; APPLICANT: Morakinyo, Layo O.
; APPLICANT: Odell, Joan T.

```

```

APPLICANT: Orozco Jr., Emil M.
APPLICANT: Rafaleki, J. Antoni
TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS
FILE REFERENCE: BB164 US NA
CURRENT APPLICATION NUMBER: US/09/659,266A
CURRENT FILING DATE: 2000-10-27
PRIOR APPLICATION NUMBER: PCT/US99/08385
PRIOR FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/083,212
PRIOR FILING DATE: 1998-04-27
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 921
TYPE: PRT
ORGANISM: Zea mays
US-09-699-266A-9

Query Match      19.3%; Score 783; DB 2; Length 921;
Best Local Similarity 27.0%; Fred. No. 2,9e-59;
Matches 251; Conservative 119; Mismatches 253; Indels 308; Gaps 411

16 RKEVKEEYMRLEQLKRFAPRADEVKSMFSSNKOILERFELLINQEMKORRIQIOPVHILTVS 75
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
49 KKRIITND--RLTYIK--NRIGENKNISYQ---RYVNS---KORQISTSKGTSAS 97
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::

76 SLNGTBE---CSVTSDDLDF-----PQVY--PLKTL 101D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
98 NLLTKQDDALCTLHS-IDLIIPVDKGGTFODESPSSNVMFGANLGPKNALIRPK-L 155D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::

102 NAVASVPIWMSNPLOQN-FWVEDETVLNIIPYMGDEVLDQGTFFIBELLKNYDGKVGCD 160D
156 PEVPKLPPTTWTFIDRRNORMEDQSVLGR-----RRIYD 191D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::

161 BECGFINDLFEVLVNALGQVNDDDDDDDGDDPERRERKQKDLBDH--RDKESRPRK 217D
192 TSGG---BALI-----CSSSEDAIEDEBEKKEFGHSDHIIIRMTVQEGC--- 233D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::

218 FPDGKIFBAISSMFPDQGAELKKEKYELFEQQLPGLAP-----ECTPNIDG 266D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
234 -MDDAVLQTLARM--ERAAADIKARFELIHGEKTKDSCKKGTENHNVKVEDLYCDKDLA 290D
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::

```



Query Match 17.9%; Score 726; DB 2; Length 689;  
Best Local Similarity 27.3%; Pred. No. 1.8e-54;  
Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

Query 267 PNAKSVOREQSLSHFTLPCR-----RCFYDCFLH-----PFAHTNTYKRNKTETA 314  
DB 221 -----ADSDNLFRCRPRQRCLVPDCKLHGCSDLVLP-----TKQAPMSGV 334  
QY 315 LDNKPCCGPOCYOHLGAKAFPAALTAER-----IKTPP-----KRPGRRR-----GRL 358  
DB 335 DDSVFCGHCHKL-----ASEPDAAGADHMLFVBEPTHSUDVMNQPSNRKKNSSGK 391  
QY 359 PNNRSRPTPTINTVL-ESKDTD-----REAGTETGENDKE-----CQTPIMKENIPEPV 430  
DB 392 TKSQSSESSSTARVISESSSEVHPISNKSPOHSPSPSKVKIIPKGIKRTINRAIERI 451  
QY 380 -----SD-----REAGTETGENDKE-----EERKQDET 404  
DB 452 LMSVKKQREMASSSNFVSGYLARDMKLNSDTRNGKRELIVSSQSSPSTRSKKST 511  
QY 405 -----SSSSSEANSR-----CQTPIMKENIPEPV 430  
DB 512 PQIGNSAPFAEANDSTEEANNRHSATDGYDSRKEBFVNEMLCKQEVYLR----- 562  
QY 431 EWSGAPASMFVLIQTYDNFCAIAR--LIGTKCRQYEFYEVKESIIAPAPABVD-- 486  
DB 563 SWKALEQGLVKGLEIFGRNSCLIAINLIGEMTKCKDVQYNNYIENNSASGALSGVSL 622  
QY 487 -----TPPRKKKKRHR-----LMAA--HCKRIQKDGSSNHYNNYPCDHP 526  
DB 623 VKGYIKGTLEKTRSYFRPRGRKVRRLKTYWSAGYNFKRTTERKQPCR--QYNPCG-C 678  
QY 527 ROPCDSSCPVIAQNFCEKFCOCSSSECONRFPQCRC-KAOCNTKQCCPYLAVERCDPDL 585  
DB 679 QSTGCKQPCLSNGCCCEKFCYCCPKICNRFPGCHCAISQGRQCPGADRECDPVC 738  
QY 586 LTC-GAAD-----HMSKSNVSCNKCSTIQSGSKHLLAPBDVAGWGIPTDQKNEF 637  
DB 729 RNCWGCGBGTGIVGNORGDNYECNNMKLLKQOQRVILGRSDVSGWGFILNKSYSKHY 798  
QY 638 ISEYCGEIIISODEADRGKVYDKWMCSPLENNPQVVDATKGNKIRPANSVAPNCA 697  
DB 799 LGETYGBELISHKAPKRGKTYDRENSSFLENINNYVLDAYTMGDILAFANHAPDNCYA 858  
QY 698 KVMVNGDHRIIGIPAKRAIQTGEELFPDYRY 728  
DB 859 KVIWVTGDHRVIGIFAKERLAGEELFPDYRY 889

RESULT 7  
US-09-177-249-2  
Sequence 2, Application US/09177249  
Patent No. 6229064  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Ohad, Nir  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadegari, Ramit  
APPLICANT: Margossian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/177,249  
CURRENT FILING DATE: 1998-10-22  
EARLIER APPLICATION NUMBER: US 09/071,838  
EARLIER FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 689  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-09-177-249-2

Query Match 17.9%; Score 726; DB 2; Length 689;  
Best Local Similarity 27.3%; Pred. No. 1.8e-54;  
Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

Query 120 FWEDEVYLANI PYWGEVLDQDGTFI---BELIKNDGKHGDEGCFINDEIFVELV 175  
DB 114 FLDEVDPLPSVKLPIYVEKLPRSLTWVFTKSQMAESD-SVIGRQIYIYNGE----- 166  
QY 176 NALGOYNDDDDDDDGDPPEF-REKQKDLJEDHRDKESRPRKPFSDIIFEAISMFPDK 234  
DB 167 -ALEISSEDEDEDEDEDEEIKKEKCEPSE-----VDRPIWVGQDY 208  
QY 225 G-----TAEHLKEKYELETQQLPGALPPECTPNIDGPNAKSVOREQSL 278  
DB 209 GLDDLIVRALAKYLKVDVSDILERYNELKAKN-----DGTAGEA--SDLTS 253  
QY 279 HSFTLF-----CRRCFYDCFLHFPFAHTNTYKRNKTETA--DNKPCGPOCYOH 327  
DB 254 KTTTAFQDPADRRRCRCMIFDCMHEKY-EPESRSESDKSSLPEDBDQPCSHCYLK 312  
QY 328 LEGAKFPAAULTAERIKTPPRKPGRRGRRLPNNSSRPTPTINTVLSEKOTDSDEAGTE 387  
DB 313 VRSVTE-----ADHV-----NDNDS----- 328  
QY 388 TGENNDKEBEKQDETSSSEANSRCQTPIMKENIPEPVENVSGAASMFVLIQTY 447  
DB 329 -----ISNKIVSDPNNTMTPEKDLIYLGILIF 358  
QY 448 YDNFCAIAR--LIGTKCRQYEFYEVKESII-----IAPAPABVD 486  
DB 359 GRNSCDVLANLIRGLKLTLEYNNVRBDDQCTMSLDLNTKTOBHNOYTKVSRKSSRSV- 417  
QY 487 TPRKKRKHLLMAHCKRIQKDGSSNHYNNYPCDHPROPCDSSCPVIAQNFCEKFC 546  
DB 418 -----RKSRLKRYAPYPALPKTTSGBKAFYKHTPTCT-CXSKGQQCPCLTHENCCKRY 472  
QY 547 CQCSSECONRFPQCRC-KAOCNTKQCCPYLAVERCDPDL-----LTCGAADHMSK-NVS 600  
DB 473 CGCSKDCNNRFGGCNCAIGQCTNRQCPFAANRECDPDLCSPLSCDGTIGETFPVQIQ 532  
QY 601 CKNCSIQSGSKHLLAPBDVAGWGIPTDQKNEFISEYCGEIIISODEADRGKVYDK 660  
DB 533 CKNQGFLLQTNKILLIGSDVHGWGAFTWDSLKQNEVIGETGBELITDBANRGRIDR 592  
QY 661 YMCSEFLNANDPVVDATRKGNKIRPANSVAPNCAVVMVNGDHRIIGIPAKRAIQTGE 720  
DB 593 IGSSYLFPLNDQLIEDARKKNEBFILNHSARPNCYAKIMTVRGDRIGLFAERAIIEGSE 652  
QY 721 ELFPDYRYSQADA 733  
DB 653 ELFPDYCYGPERA 665

RESULT 8  
US-09-061-769A-2  
Sequence 2, Application US/09061769A  
Patent No. 6239327  
GENERAL INFORMATION:  
APPLICANT: Cold Spring Harbor Labs  
TITLE OF INVENTION: No. 6239327el Seed Specific Polycomb Group Gene and Methods of  
TITLE OF INVENTION: use for Same  
FILE REFERENCE: mea genes  
CURRENT APPLICATION NUMBER: US/09/061,769A  
CURRENT FILING DATE: 1998-04-16  
PRIOR APPLICATION NUMBER: 09/061,769  
PRIOR FILING DATE: 1998-04-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 689  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: MEAMT  
US-09-061-769A-2

17.9%; Score 726; DB 2; Length 689;

Query Match Best Local Similarity 27.3%; Pred. No. 1.8e-54; Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

```

Qy 120 FMVEDETVLANIPFMGDEVLDDQDTFI-----EELIKNYDGKVGHDREGCFINDEIFVELV 175
Db 114 FLDEVDLPLPSVKLPIVEKLPRLSITWFTKSSQLMAESD-SVIGKQIYLLNGE----- 166
Qy 176 NALGQYNDDDDDGDDPEE-REBKQDLBDHDDKESRPRKFPDSKIFEAISMPDPK 234
Db 167 -ALBLSSEBDEBDEBEIKKEKCEFSBD-----VDRFIWTVGQDY 208
Qy 235 G-----TAELEKKEKYELETEQOLPGALPPECTPNIDGNPAKSVQREOSL 278
Db 209 GLDDLVRRALAKYLEVDVSDILERYNELKLN-----DGTAGEA--SDLTS 253
Qy 279 HSFHTLF-----CRCKFYDCFLPHFATPTTYKKTETAL---DNKPCGPOCYQH 327
Db 254 KITTAQDPADRRHRCRCMIFDCHMEKY-EPESRSEBKSILFEDEDRQPCSEHCYLK 312
Qy 328 LEGAKEFAALTAERIKTPPRKPGRRGRGLPNNSRPSTPTINVLSEKDTDSREAGTE 387
Db 313 VRSVTE-----ADHV-----MDNDS----- 328
Qy 388 TGGENDKEBEKKDETSSSEANSRCQTPIKMKPNIEPBNVEMSGAEMFVILGTY 447
Db 329 -----ISNKLIVSDPNNTMTTVEKOLYLKGEIF 358
Qy 448 YDNFCALAR--LIGTKCRQYEFERVKESI-----IAPAPABDVD 486
Db 359 GRNSGDVALNLRGLKCLLEYNWRBDDQCTMSLDLNTKTTQRHNQYTKVSRKSSRSV- 417
Qy 487 TPRPKKGRKRLMAAHCRIQLKKDGSNNHYNYOPCDHPRQCDSSCPVIAQNFCEKF 546
Db 418 ---RKSRLKRYAYPPALKKTTSGBAKFYKHYPCT-CRSKCGQCPCLTHNCCCKY 472
Qy 547 CQGSSECONRPPGCRCK-AQCNTKQPCYLAVRECDPDLG---LTGAADHWDK-NVS 600
Db 473 CGCSKDCNNRFGGCGNCAIGQCTNRCPCFPAANRECDPDLGSCPLSCGDGTGLETVPVQIQ 532
Qy 601 CKNGSIGRSGKHLILAPSDVAGWGIPTKDPVQKNEIFSEYCGEIIISODEADBRGKYVDK 660
Db 533 CKNNQFLQTNKKLILIGKSDVHGWAFTWDSLKKNBYGETYGLITLHDEANERGRIDR 592
Qy 661 YMCSEFLNANDPVVDATRKGNKIRFANHSVNPNCYAVMMVNGDHRIGIFAKRAIQTGE 720
Db 593 IGSSYLFLLANDQLEIDARKKNEFPFLNHSARFNCYAKLMTVRDQRIGLFAERAIERGE 652
Qy 721 ELFPDYRYSQADA 733
Db 653 ELFPDYCYGPEHA 665

```

RESULT 9  
US-09-812-283-2  
Sequence 2, Application US/09812283  
Patent No. 6828477  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Ohad, Nir  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadegari, Ramtin  
APPLICANT: Margosian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/812.283

CURRENT FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 09/177,249

PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: US 09/071,838

PRIOR FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2  
LENGTH: 689  
TYPE: PRT

ORGANISM: Arabidopsis sp.  
US-09-812-283-2

17.9%; Score 726; DB 2; Length 689;

Query Match Best Local Similarity 27.3%; Pred. No. 1.8e-54; Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

```

Qy 120 FMVEDETVLANIPFMGDEVLDDQDTFI-----EELIKNYDGKVGHDREGCFINDEIFVELV 175
Db 114 FLDEVDLPLPSVKLPIVEKLPRLSITWFTKSSQLMAESD-SVIGKQIYLLNGE----- 166
Qy 176 NALGQYNDDDDDGDDPEE-REBKQDLBDHDDKESRPRKFPDSKIFEAISMPDPK 234
Db 167 -ALBLSSEBDEBDEBEIKKEKCEFSBD-----VDRFIWTVGQDY 208
Qy 235 G-----TAELEKKEKYELETEQOLPGALPPECTPNIDGNPAKSVQREOSL 278
Db 209 GLDDLVRRALAKYLEVDVSDILERYNELKLN-----DGTAGEA--SDLTS 253
Qy 279 HSFHTLF-----CRCKFYDCFLPHFATPTTYKKTETAL---DNKPCGPOCYQH 327
Db 254 KITTAQDPADRRHRCRCMIFDCHMEKY-EPESRSEBKSILFEDEDRQPCSEHCYLK 312
Qy 328 LEGAKEFAALTAERIKTPPRKPGRRGRGLPNNSRPSTPTINVLSEKDTDSREAGTE 387
Db 313 VRSVTE-----ADHV-----MDNDS----- 328
Qy 388 TGGENDKEBEKKDETSSSEANSRCQTPIKMKPNIEPBNVEMSGAEMFVILGTY 447
Db 329 -----ISNKLIVSDPNNTMTTVEKOLYLKGEIF 358
Qy 448 YDNFCALAR--LIGTKCRQYEFERVKESI-----IAPAPABDVD 486
Db 359 GRNSGDVALNLRGLKCLLEYNWRBDDQCTMSLDLNTKTTQRHNQYTKVSRKSSRSV- 417
Qy 487 TPRPKKGRKRLMAAHCRIQLKKDGSNNHYNYOPCDHPRQCDSSCPVIAQNFCEKF 546
Db 418 ---RKSRLKRYAYPPALKKTTSGBAKFYKHYPCT-CRSKCGQCPCLTHNCCCKY 472
Qy 547 CQGSSECONRPPGCRCK-AQCNTKQPCYLAVRECDPDLG---LTGAADHWDK-NVS 600
Db 473 CGCSKDCNNRFGGCGNCAIGQCTNRCPCFPAANRECDPDLGSCPLSCGDGTGLETVPVQIQ 532
Qy 601 CKNGSIGRSGKHLILAPSDVAGWGIPTKDPVQKNEIFSEYCGEIIISODEADBRGKYVDK 660
Db 533 CKNNQFLQTNKKLILIGKSDVHGWAFTWDSLKKNBYGETYGLITLHDEANERGRIDR 592
Qy 661 YMCSEFLNANDPVVDATRKGNKIRFANHSVNPNCYAVMMVNGDHRIGIFAKRAIQTGE 720
Db 593 IGSSYLFLLANDQLEIDARKKNEFPFLNHSARFNCYAKLMTVRDQRIGLFAERAIERGE 652
Qy 721 ELFPDYRYSQADA 733
Db 653 ELFPDYCYGPEHA 665

```

RESULT 10  
US-10-176-884-14  
Sequence 14, Application US/10176884  
Patent No. 6906244  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert  
APPLICANT: Kinoshita, Tetsu

APPLICANT: Yadegari, Ramin  
 APPLICANT: Gehring, Mary  
 APPLICANT: Okamoto, Jack  
 APPLICANT: Dang, Van-Dinh  
 APPLICANT: The Regents of the University of California  
 APPLICANT: Ceres, Inc.  
 TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
 TITLE OF INVENTION: Development  
 FILE REFERENCE: 023070-116710US  
 CURRENT APPLICATION NUMBER: US/10/176,884  
 CURRENT FILING DATE: 2002-06-21  
 PRIOR APPLICATION NUMBER: US 60/300,506  
 PRIOR FILING DATE: 2001-06-22  
 NUMBER OF SEQ ID NOS: 91  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 14  
 LENGTH: 689  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 FEATURE:  
 OTHER INFORMATION: MEDA (MEA)  
 US-10-176-884-14

Query Match 17.9%; Score 726; DB 2; Length 689;  
 Best Local Similarity 27.3%; Pred. No. 1,8e-54;  
 Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

120 FMVEDETVHNIPYMGDEVLDQDGTFI---BELIKYDGVHGDRCGFINDLIFVELV 175  
 114 FLDEVDPLPSVKLPYVEKLPRSTIWTFTKSSQLMAESD-SVIGRQIYYLNGE----- 166  
 176 NALGOYNDDDDDDDGDPPEE-REEKOKLEDRDKESRPPKPSDKIFRAISSMPDK 234  
 167 -ALBLSSEDEDEDEDEBEIKKCEPSED-----VDREPIWTVGQDY 208  
 235 G-----TABELKERYKELTEQQLPGALPPECTPNIIDGPNKSVQREOSL 278  
 209 GLDILVVRALAKYLEVDVSDILERYNELKLN-----DTAGEA--SDILTS 253  
 279 HSFHTLF-----CRRCFYDCFLPHFATPNTYKRYKNTETAL---DNKPCGPOCYOH 327  
 254 KITTAFOQDPADRHRCHRCMTFDCMHKRY-EPERSSEDSLSLPEDDRQPCSHCYLK 312  
 328 LEGAKEFAALTAERIKTPPRKPGRRGRRLPNNSRSPSTPTINVLBSKOTDSREACTE 387  
 313 VRSVTE-----ADHV-----MDNDS----- 328  
 388 TGGENNDEBEEKOETSSSSSEANSRCQPTIMKKNINIPPENVENSGAASMFVLYIGTY 447  
 329 -----ISNKIVSDPENTMTPTVEKDIYKGIET 358  
 448 YDNFCALAR--LIGTKCROVEFRVKESSI-----IAPAEADVD 486  
 359 GRNSCDVALNLRGLKTCLEIYNWRBEDQCTMSLDLNTKQTHNQVTKKYSRKSRSRY- 417  
 487 TPRKRRKRLMAAHCRTKIQKDGSSNHYVNTYQPCHPROPDSSGCPYIAQNFCEKY 546  
 418 ---RKSRLRYAYPPALKTKTTSGEAKFYKHYPCT-CKSKCCQCPCLHNECKERY 472  
 547 CGCSECCNRPFGCGCK-AQCNTKQCPYLAVRECDPLC-----LTGAAADHMDSK-NVS 600  
 473 CGCSDCCNRRFGCGCAIGQCTNRQCPANRECDPLCRSCPLSCGDTLGETPPVQIQ 532  
 601 CKNCISGRGSKHLLAPSDVAGWGIPTKDPVOKNEFISEYCGEILISODEADRGRKYVDK 660  
 533 CKNMQFLQTNKKLIIIGKSDVAGWGAFTMDSLKKNVEYIGEYIGELITDEANERIRIDR 592  
 661 YMCSTLFUNLNDFFVDATRKGNKIRFPAHNSVNPNCYAYMMVNGDHRIGIFAKRAIQTGE 720  
 593 IGSSTLFTLNQOLEIDARRKGNFPEFLNHSAPNCTAKLMTIVRGDRIGLFAERAIIBEGE 652  
 721 ELFPDYRSQADA 733

DB 653 ELFPDYCYGPEHA 665

RESULT 11

US-09-071-838A-2  
 Sequence 2, Application US/09071838A  
 Patent No. 7029917

GENERAL INFORMATION:

APPLICANT: Fischer, Robert L.

Onad, Nir

Kiyosue, Tomohiro

Yadegari, Ramon

Margosian, Linda

Harada, John

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Goldberg, Robert B.

Query Match 17.9%; Score 726; DB 3; Length 689;

Best Local Similarity 27.3%; Pred. No. 1,8e-54;

Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;

120 FMVEDETVHNIPYMGDEVLDQDGTFI---BELIKYDGVHGDRCGFINDLIFVELV 175  
 114 FLDEVDPLPSVKLPYVEKLPRSTIWTFTKSSQLMAESD-SVIGRQIYYLNGE----- 166  
 176 NALGOYNDDDDDDDGDPPEE-REEKOKLEDRDKESRPPKPSDKIFRAISSMPDK 234  
 167 -ALBLSSEDEDEDEDEBEIKKCEPSED-----VDREPIWTVGQDY 208  
 235 G-----TABELKERYKELTEQQLPGALPPECTPNIIDGPNKSVQREOSL 278  
 209 GLDILVVRALAKYLEVDVSDILERYNELKLN-----DTAGEA--SDILTS 253  
 279 HSFHTLF-----CRRCFYDCFLPHFATPNTYKRYKNTETAL---DNKPCGPOCYOH 327  
 254 KITTAFOQDPADRHRCHRCMTFDCMHKRY-EPERSSEDSLSLPEDDRQPCSHCYLK 312  
 328 LEGAKEFAALTAERIKTPPRKPGRRGRRLPNNSRSPSTPTINVLBSKOTDSREACTE 387  
 313 VRSVTE-----ADHV-----MDNDS----- 328

QY 388 TGGENNDEKEEKKDETTSSSSBANSRCOTPIKMKPNIEPPENYEWGASAMFVLLIGTY 447  
Db 329 -----ISNKTIIVSDPNNTMTWTFVEKOLYLKGIEIF 358  
QY 448 YDNFCALAR--LIGTKTCROYEPRVYESSI-----IAPAPADVD 486  
Db 359 GRNSCDVALNIIIRGLKTCLEIYNYMRBODQCTWSLDINKTTOHRNQYTKVSRKSSRV- 417  
QY 487 TPRKKKKRKRRLMAAHCCKIQLKKDGSNNHYNYOPCDHROPDCSSCPVIAONFCEKF 546  
Db 418 ----RKSRLRKRYARYPALAKKTTSGSAKPYKHTTPT-CKSKGQOQCPCLTHENCCEKY 472  
QY 547 CQCSSECONRPPGCRCK-AQCNTRQCPCYLAVERCDPLC---LTCGAADHMSK-NVS 600  
Db 473 CGSKDCNNRPPGGNCALIGQCTNRQCPFANRECDPLCGSCPLSGDGTIGETPVQIQ 532  
QY 601 CKKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCEIISQDEADRGRKYDK 660  
Db 533 CKKMQFLQNTKKILIGKSDVHGWAFTWDSLKKNBYLGEYTGELITHDEANERGRJEDR 592  
QY 661 YMCSEFLNANDPVVDATRKGNKIRIPANHSVNPNCYAKVMVNGDHRIGIFAKRALOTGE 720  
Db 593 IGSSYLTTLNDQLEIDARRKGNBEKFLNHSARPNCYAKMLIVRGDQRIGLFAERAIIEGE 652  
QY 721 ELFPDYRYSOADA 733  
Db 653 ELFPDYCYGPEHA 665  
RESULT 12  
US-10-176-884-15  
Sequence 15, Application US/10176884  
Patent No. 6906244  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert  
APPLICANT: Kinoshita, Tetsu  
APPLICANT: Yadegari, Ramlin  
APPLICANT: Gehring, Mary  
APPLICANT: Okamuro, Jack  
APPLICANT: Dang, Van-Dinh  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
FILE REFERENCE: 023070-116710US  
CURRENT APPLICATION NUMBER: US/10/176,884  
PRIOR FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/300,506  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 15  
LENGTH: 932  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: MEA-15g-ESAI  
OTHER INFORMATION: chimeric polypeptide encoded by transgene plasmid  
OTHER INFORMATION: pCR304-5UAS MEA-15g-ESAI  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (906)  
OTHER INFORMATION: Xaa = Gly or Glu  
US-10-176-884-15  
Query Match 17.9%; Score 726; DB 2; Length 932;  
Best local similarity 27.3%; Pred. No. 2.8e-54;  
Matches 184; Conservative 104; Mismatches 205; Indels 180; Gaps 21;  
QY 120 FMVDEDTYLNIPMGDVLADODGTFI---EELIKVNDGVHGDREGCFINDSIDFVFLV 175  
Db 114 FLDEDDVPLPSVKLPYVKELPRSTITWTFKSSQLMAESD-SVIGKQIYLYNGE----- 166

QY 176 NALGQYNDDDDDGDDPPEE-REEKQDLDEHDDXESRPKPFESDKIFEAISMPEDK 224  
Db 167 -ALELSEDEDEDEDEEIKKEKCEFSH-----VDRAFVWGQDY 208  
QY 235 G-----TAELEKRYKELTEQQLPGALPPECTPNIDGPAAKSVQREOSL 278  
Db 209 GLDDLVRALAKYLEVDVSDILERYNELTKN-----DGTAGEA--SDLTIS 253  
QY 279 HSFHTLF-----GRCEKYDCFLHPFATENTYKRNKTERAL---DNKPGPOCYOH 327  
Db 254 KTTTFNFOADRRHRCRCMI FDCMHBEKX-EPESRSBEDKSLFDEDEDQPCSEHCYLK 312  
QY 328 LBEAKFPAALTYERIKTPPKRPGRRRGGLPNSSRPSTPTTINVLSEKDTSDREAGTE 387  
Db 313 VRSVTE-----ADHV-----MDNDS----- 328  
QY 388 TGGENNDEKEEKKDETTSSSSBANSRCQPIKMKPNIEPPENYEWGASAMFVLLIGTY 447  
Db 329 -----ISNKTIIVSDPNNTMTWTFVEKOLYLKGIEIF 358  
QY 448 YDNFCALAR--LIGTKTCROYEPRVYESSI-----IAPAPADVD 486  
Db 359 GRNSCDVALNIIIRGLKTCLEIYNYMRBODQCTWSLDINKTTOHRNQYTKVSRKSSRV- 417  
QY 487 TPRKKKKRKRRLMAAHCCKIQLKKDGSNNHYNYOPCDHROPDCSSCPVIAONFCEKF 546  
Db 418 ----RKSRLRKRYARYPALAKKTTSGSAKPYKHTTPT-CKSKGQOQCPCLTHENCCEKY 472  
QY 547 CQCSSECONRPPGCRCK-AQCNTRQCPCYLAVERCDPLC---LTCGAADHMSK-NVS 600  
Db 473 CGSKDCNNRPPGGNCALIGQCTNRQCPFANRECDPLCGSCPLSGDGTIGETPVQIQ 532  
QY 601 CKKCSIORGSKKHLILAPSDVAGWGIPIKDPVQKNEFISEYCEIISQDEADRGRKYDK 660  
Db 533 CKKMQFLQNTKKILIGKSDVHGWAFTWDSLKKNBYLGEYTGELITHDEANERGRJEDR 592  
QY 661 YMCSEFLNANDPVVDATRKGNKIRIPANHSVNPNCYAKVMVNGDHRIGIFAKRALOTGE 720  
Db 593 IGSSYLTTLNDQLEIDARRKGNBEKFLNHSARPNCYAKMLIVRGDQRIGLFAERAIIEGE 652  
QY 721 ELFPDYRYSOADA 733  
Db 653 ELFPDYCYGPEHA 665  
RESULT 13  
US-09-699-266A-11  
Sequence 11, Application US/09699266A  
Patent No. 6559354  
GENERAL INFORMATION:  
APPLICANT: Ma, Hongchang  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Odell, Joan T.  
APPLICANT: Orozco Jr., Emil M.  
APPLICANT: Rafalski, J. Antoni  
TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS  
FILE REFERENCE: BB1164 US NA  
CURRENT APPLICATION NUMBER: US/09/699,266A  
PRIOR FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: PCT/US99/08385  
PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/083,212  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 11  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Trifolium aestivum  
US-09-699-266A-11  
Query Match 14.2%; Score 576.5; DB 2; Length 359;

Best Local Similarity 38.3%; Pred. No. 7.9e-42;  
Matches 128; Conservative 44; Mismatches 129; Indels 33; Gaps 7;

QY 432 MSGAASRFRVLIGTYTDFCAIAR--LIGTCTCKOV--YERPRVSSSIAPAPBDVT 487  
DB 3 WSLTRDLYLKQIEIFGKNSCLIVNLLCGLTCTCHEVASMYNMGANNSKSIQDPFET 62  
QY 488 PE-----RKKGRRLMAAHCRTLOKKQSSNHYVNPQCDHPP 529  
DB 63 EONTNBOGVVTKYCRRRGRTRKTKYPSKAGHPAIRKKVGDGKQCDRQTPCG--CQEM 121  
QY 530 CDSGPCVLAQNFCEKFCQSSSECONRPFGRCC--XAQNTKQCPCLAVRECDPDLCLTC 588  
DB 122 CMKNCPCVENGTCCCEYGCSCSKNRRFGCHCAKISAGANAHVLLPBGVNIYMFGRNC 181  
QY 589 -----GAADHWSKXV--SCNCSIQSGSKKHLILAPSDVAGNGIFIKDPVQXNRFIS 639  
DB 182 WGLWGSFRRATYKKEVMTVQCGNMKLLKQQRILGKSDVAGWGAFTKNGVHKNDYLG 241  
QY 640 EYCGEIIISQDEADRRGKVVDKVMCSFLPYLNDPVDATRKNGKIRFANHSVNPVCYAKV 699  
DB 242 EYTGELISHKRDADKKGKIDRANSSFLPDLNDQFVLDAVRKDKLKFANHSSPVCYAKV 301  
QY 700 MMVNGDHRIGIFAKRAIQTEBELFPDYRYSQADA 733  
DB 302 MMVAGDHRVGIYAREHIEBSALEFYDYRYPDOA 335

RESULT 14  
US-10-200-012-39  
; Sequence 39, Application US/10200012  
; Patent No. 6955905  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: PR/SET- Domain Containing Nucleic Acids,  
; TITLE OF INVENTION: Polypeptides, Antibodies and Methods of Use  
; FILE REFERENCE: P-LJ 5301  
; CURRENT APPLICATION NUMBER: US/10/200,012  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: US 09/910,478  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 114  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,  
; LOCATION: 46, 47, 48, 49, 50, 51  
; OTHER INFORMATION: synthetic peptide  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-200-012-39

Query Match 11.7%; Score 474; DB 2; Length 114;  
Best Local Similarity 81.6%; Pred. No. 1.4e-33;  
Matches 93; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 621 VAGMGIFIKDPVQXNRFISEYCGEIIISQDEADRRGKVVDKVMCSFLPYLNDPVDATRK 680  
DB 1 VAGMGIFIKDPVQXNRFISEYCGEIIISQDEKXXXXXXXXXXXXXXXXXADPVDATRK 60  
QY 681 GNKIRFANHSVNPVCYAKVMVNGDHRIGIFAKRAIQTEBELFPDYRYSQADAL 734  
DB 61 GNKIRFANHSVNPVCYAKVMVNGDHRIGIFAKRAIQTEBELFPDYRYSQADAL 114

RESULT 15  
US-10-176-884-19  
; Sequence 19, Application US/10176884  
; Patent No. 6906244  
; GENERAL INFORMATION:

; APPLICANT: Fischer, Robert  
; APPLICANT: Kinoshita, Tetsu  
; APPLICANT: Yadegari, Ramon  
; APPLICANT: Gehring, Mary  
; APPLICANT: Okamoto, Jack  
; APPLICANT: Dang, Van-Dinh  
; APPLICANT: The Regents of the University of California  
; APPLICANT: Ceres, Inc.  
; TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
; FILE REFERENCE: 023070-116710US  
; CURRENT APPLICATION NUMBER: US/10/176,884  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/300,506  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 115  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:exemplary SET  
; OTHER INFORMATION: domain  
US-10-176-884-19

Query Match 9.0%; Score 365; DB 2; Length 115;  
Best Local Similarity 56.5%; Pred. No. 4.5e-24;  
Matches 65; Conservative 20; Mismatches 30; Indels 0; Gaps 0;

QY 619 SDVAGMGIFIKDPVQXNRFISEYCGEIIISQDEADRRGKVVDKVMCSFLPYLNDPVDAT 678  
DB 1 SDVAGMGIFIKDPVQXNRFISEYCGEIIISQDEADRRGKVVDKVMCSFLPYLNDPVDAT 60  
QY 679 RKGKIRFANHSVNPVCYAKVMVNGDHRIGIFAKRAIQTEBELFPDYRYSQADA 733  
DB 61 RKGKIRFANHSVNPVCYAKVMVNGDHRIGIFAKRAIQTEBELFPDYRYSQADA 115

Search completed: August 11, 2006, 22:30:20  
Job time : 42.1709 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:45:12 ; Search time 127.001 Seconds  
(without alignments)  
2720.917 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGQTGKSEKPEVCWKRRKRVK.....RYSQADAKYVIEREMKIP 746

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4051	100.0	746	5	US-10-773-302-2
2	4038	99.7	746	4	US-10-153-668-232
3	4038	99.7	746	4	US-10-231-778-225
4	4038	99.7	746	4	US-10-295-027-710
5	3993	98.6	746	4	US-10-231-778-226
6	3812.5	94.1	707	4	US-10-104-047-3162
7	3812.5	94.1	707	6	US-11-072-512-3162
8	2845	70.2	517	4	US-10-153-668-354
9	2429.5	60.0	723	4	US-10-115-482-56
10	2156.5	53.2	760	6	US-11-097-143-14073
11	2149.5	55.1	760	4	US-10-231-778-224
12	2131.5	52.6	760	5	US-10-773-302-11
13	1208.5	29.8	242	4	US-10-424-599-219717
14	829	20.5	895	5	US-10-739-930-7656
15	828.5	20.5	856	4	US-10-231-778-221
16	828.5	20.5	880	4	US-10-425-114-72833
17	812	20.0	685	4	US-10-424-599-174013
18	810	20.0	893	3	US-09-906-453-4
19	810	20.0	893	6	US-11-230-145-4
20	805	19.9	891	4	US-10-310-154-599
21	795	19.6	931	4	US-10-425-115-281072
22	795	19.6	931	5	US-10-739-930-7340
23	785.5	19.4	1015	4	US-10-437-963-105693
24	784	19.4	895	4	US-10-437-963-146048
25	782.5	19.3	902	4	US-10-231-778-222
26	781	19.3	931	3	US-09-906-453-2
27	781	19.3	931	6	US-11-230-145-2

28	777	19.2	699	4	US-10-437-963-146070	Sequence 146070,
29	775.5	19.1	958	4	US-10-425-115-281076	Sequence 281076,
30	769	19.0	895	5	US-10-732-923-15015	Sequence 15015, A
31	731	18.0	509	4	US-10-425-114-57456	Sequence 57456, A
32	726	17.9	689	3	US-09-071-838-2	Sequence 2, Appl1
33	726	17.9	689	4	US-10-213-512-2	Sequence 2, Appl1
34	726	17.9	689	4	US-10-176-884-14	Sequence 14, Appl1
35	726	17.9	689	4	US-10-231-778-1	Sequence 1, Appl1
36	726	17.9	689	4	US-10-177-478-5	Sequence 5, Appl1
37	726	17.9	689	6	US-11-051-456-14	Sequence 14, Appl1
38	726	17.9	932	4	US-10-176-884-15	Sequence 15, Appl1
39	726	17.9	932	4	US-10-177-478-21	Sequence 21, Appl1
40	726	17.9	932	6	US-11-051-456-15	Sequence 15, Appl1
41	719.5	17.8	463	4	US-10-425-114-57452	Sequence 57452, A
42	713	17.6	417	4	US-10-425-114-57969	Sequence 57969, A
43	700.5	17.3	536	4	US-10-425-114-69287	Sequence 69287, A
44	681.5	16.8	346	4	US-10-425-114-57460	Sequence 57460, A
45	659	16.3	294	4	US-10-425-114-69819	Sequence 69819, A

## ALIGNMENTS

RESULT 1  
US-10-773-302-2

Sequence 2, Application US/10773302

Publicatation No. US2005008980A1  
GENERAL INFORMATION:

APPLICANT: Jenuwein, Thomas

APPLICANT: Laible, Gotz

APPLICANT: O'Carroll, Donal

APPLICANT: Eisenhaber, Frank

APPLICANT: Rea, Stephen

TITLE OF INVENTION: Chromatin-Regulator Genes

FILE REFERENCE: 0652.1670001

CURRENT APPLICATION NUMBER: US/10/773.302

CURRENT FILING DATE: 2004-02-09

PRIOR APPLICATION NUMBER: US/09/589.892

PRIOR FILING DATE: 2000-06-09

PRIOR APPLICATION NUMBER: US 08/945.988

PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: PCT/EP96/01818

PRIOR FILING DATE: 1996-05-02

PRIOR APPLICATION NUMBER: DE 195 16 776.7

PRIOR FILING DATE: 1995-05-10

NUMBER OF SEQ ID NOS: 21

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2

LENGTH: 746

TYPE: PRT

ORGANISM: Homo sapiens

US-10-773-302-2

Query Match 100.0%; Score 4051; DB 5; Length 746;  
Best Local Similarity 100.0%; Pred. No. 9.4e-284; Indels 0; Gaps 0;  
Matches 746; Conservative 0; Mismatches 0;

QY	1	MGQTGKSEKPEVCWKRRKRVKSEYWRQLKRRRADEVKSMFSSNRQKILRETEILNDEW	60
DB	1	MGQTGKSEKPEVCWKRRKRVKSEYWRQLKRRRADEVKSMFSSNRQKILRETEILNDEW	60
QY	61	KORRIOPVHIIITVSLSLGRFECVTSPLDPTQYIPLKTLNANAVSVIATWSPLQONF	120
DB	61	KORRIOPVHIIITVSLSLGRFECVTSPLDPTQYIPLKTLNANAVSVIATWSPLQONF	120
QY	121	MVEDETVLANIPLYMGDEVLDGDTFIEELINYGKVGDBRCGFINDEIFVELVNALGQ	180
DB	121	MVEDETVLANIPLYMGDEVLDGDTFIEELINYGKVGDBRCGFINDEIFVELVNALGQ	180
QY	181	YNDDDDDDDGDDPESREKQKDLSDHRDDEKSRPPRKPSDKITFEALISMPDPKGTABEL	240
DB	181	YNDDDDDDDGDDPESREKQKDLSDHRDDEKSRPPRKPSDKITFEALISMPDPKGTABEL	240

QY 241 KEKTELTEQOLPALPPECTPNTIDGNPAKSVORBSLHSHFTLFCRRCFYTCFLHPH 300  
| | | | |  
Db 241 KEKTELTEQOLPALPPECTPNTIDGNPAKSVORBSLHSHFTLFCRRCFYTCFLHPH 300  
QY 301 ATPVTYRKNTETALDNKPCGPOCYOHLGAKGFPAALTAERIKTPPKRPGRRGRLPN 360  
| | | | |  
Db 301 ATPVTYRKNTETALDNKPCGPOCYOHLGAKGFPAALTAERIKTPPKRPGRRGRLPN 360  
QY 361 NSSRPTPTINVLBSKOTSDREAGTETGGENNDKEBEKKDETSSSEANSRCQTP1KM 420  
| | | | |  
Db 361 NSSRPTPTINVLBSKOTSDREAGTETGGENNDKEBEKKDETSSSEANSRCQTP1KM 420  
QY 421 KPNTPEPENWMSGASAMFVULIGTYDNFCALRIIGYTCQOYEFVYKSSIIAPA 480  
| | | | |  
Db 421 KPNTPEPENWMSGASAMFVULIGTYDNFCALRIIGYTCQOYEFVYKSSIIAPA 480  
QY 481 PAEVDTPPRKKRKHRLMAHCRKIQLKDGSSNHVNYNPOCDHPROPCDSSCPVIAQ 540  
| | | | |  
Db 481 PAEVDTPPRKKRKHRLMAHCRKIQLKDGSSNHVNYNPOCDHPROPCDSSCPVIAQ 540  
QY 541 NFCEKFCQCSSECONRPFPGCRCAQCNTRKQPCYLAVERCDPDLCTCGAADHDSKNVS 600  
| | | | |  
Db 541 NFCEKFCQCSSECONRPFPGCRCAQCNTRKQPCYLAVERCDPDLCTCGAADHDSKNVS 600  
QY 601 CKNCISIORGSKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRGRKYDK 660  
| | | | |  
Db 601 CKNCISIORGSKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRGRKYDK 660  
QY 661 YMSCFLFNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
| | | | |  
Db 661 YMSCFLFNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
QY 721 ELFPDYRYSQADALKYVGIEREMEIP 746  
| | | | |  
Db 721 ELFPDYRYSQADALKYVGIEREMEIP 746

## RESULT 2

US-10-153-668-232  
; Sequence 232, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUMA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STATe Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; PRIOR FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 232  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-153-668-232

Query Match 99.7%; Score 4038; DB 4; Length 746;  
Best Local Similarity 99.7%; Pred. No. 8,2e-283;  
Matches 744; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGOTGKSEKGPVCCWRRKRVSEYMRILQLRFRPADVYSMBSSNRKIIERTIELNQEW 60  
| | | | |  
Db 1 MGOTGKSEKGPVCCWRRKRVSEYMRILQLRFRPADVYSMBSSNRKIIERTIELNQEW 60  
QY 61 KORRIQPVHILITVSSIRGTRECSVTSDDLPPQVILPLKTLNVAAPIMYSPLQONF 120  
| | | | |  
Db 61 KORRIQPVHILITVSSIRGTRECSVTSDDLPPQVILPLKTLNVAAPIMYSPLQONF 120  
QY 121 MVEDEVIANIPLYMGDEVLDODGFIIEELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
| | | | |  
Db 121 MVEDEVIANIPLYMGDEVLDODGFIIEELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
QY 181 YNDDDDDDDDDEPEREEKODLEDRDYESSRPPKRPBQDKFEAISMPDQKTAEBL 240  
| | | | |  
Db 181 YNDDDDDDDDDEPEREEKODLEDRDYESSRPPKRPBQDKFEAISMPDQKTAEBL 240  
QY 241 KEKTELTEQOLPALPPECTPNTIDGNPAKSVORBSLHSHFTLFCRRCFYTCFLHPH 300  
| | | | |  
Db 241 KEKTELTEQOLPALPPECTPNTIDGNPAKSVORBSLHSHFTLFCRRCFYTCFLHPH 300  
QY 301 ATPVTYRKNTETALDNKPCGPOCYOHLGAKGFPAALTAERIKTPPKRPGRRGRLPN 360  
| | | | |  
Db 301 ATPVTYRKNTETALDNKPCGPOCYOHLGAKGFPAALTAERIKTPPKRPGRRGRLPN 360  
QY 361 NSSRPTPTINVLBSKOTSDREAGTETGGENNDKEBEKKDETSSSEANSRCQTP1KM 420  
| | | | |  
Db 361 NSSRPTPTINVLBSKOTSDREAGTETGGENNDKEBEKKDETSSSEANSRCQTP1KM 420  
QY 421 KPNTPEPENWMSGASAMFVULIGTYDNFCALRIIGYTCQOYEFVYKSSIIAPA 480  
| | | | |  
Db 421 KPNTPEPENWMSGASAMFVULIGTYDNFCALRIIGYTCQOYEFVYKSSIIAPA 480  
QY 481 PAEVDTPPRKKRKHRLMAHCRKIQLKDGSSNHVNYNPOCDHPROPCDSSCPVIAQ 540  
| | | | |  
Db 481 PAEVDTPPRKKRKHRLMAHCRKIQLKDGSSNHVNYNPOCDHPROPCDSSCPVIAQ 540  
QY 541 NFCEKFCQCSSECONRPFPGCRCAQCNTRKQPCYLAVERCDPDLCTCGAADHDSKNVS 600  
| | | | |  
Db 541 NFCEKFCQCSSECONRPFPGCRCAQCNTRKQPCYLAVERCDPDLCTCGAADHDSKNVS 600  
QY 601 CKNCISIORGSKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRGRKYDK 660  
| | | | |  
Db 601 CKNCISIORGSKHLLAPSDVAGWGIPTKDPVQKNEFISEYCGEIIISQDEADRGRKYDK 660  
QY 661 YMSCFLFNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
| | | | |  
Db 661 YMSCFLFNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
QY 721 ELFPDYRYSQADALKYVGIEREMEIP 746  
| | | | |  
Db 721 ELFPDYRYSQADALKYVGIEREMEIP 746

## RESULT 3

US-10-231-778-225  
; Sequence 225, Application US/10231778  
; Publication No. US20030126647A1  
; GENERAL INFORMATION:  
; APPLICANT: Biodeau, Pierre  
; APPLICANT: Chaudhury, Abdul M.  
; APPLICANT: Dennis, Elizabeth S.  
; APPLICANT: Koltunow, Anna M.G.  
; APPLICANT: Luo, Ming  
; APPLICANT: Peacock, William J.  
; TITLE OF INVENTION: Method for inducing seed development by down-regulating  
; FILE REFERENCE: 72-98A  
; CURRENT APPLICATION NUMBER: US/10/231,778  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: 09/398,237  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 60/101,184





Db 181 YNDDDDDDGDDPDEREREKQLEDRDDEKSRPPRPFPSDKLLEAISMSFPDKGTAEEL 240  
Qy 241 KEKYEKELTEOOLPGALPPECTPNIDGPNKSVOREGSLHSFHTLFCRRCFKYDCFLHPH 300  
Db 241 KEKYEKELTEOOLPGALPPECTPNIDGPNKSVOREGSLHSFHTLFCRRCFKYDCFLHPH 300  
Qy 301 ATPNTYRKKTETALDNKPCGPOCYOHLGAKEPFAALTABRIKTPPRKPGRRGRGLPN 360  
Db 301 ATPNTYRKKTETALDNKPCGPOCYOHLGAKEPFAALTABRIKTPPRKPGRRGRGLPN 360  
Qy 361 NSSRPSPTITNVLSKOTDSREAGTETGSENNDEEEKDETSSESEANSRCQTP1KM 420  
Db 361 NSSRPSPTITNVLSKOTDSREAGTETGSENNDEEEKDETSSESEANSRCQTP1KM 420  
Qy 421 KPNIEPPEVMSGAEASMPFVLIGTYINDFCALRLIGTKCQVYEFKVESSTIAPA 480  
Db 421 KPNIEPPEVMSGAEASMPFVLIGTYINDFCALRLIGTKCQVYEFKVESSTIAPA 480  
Qy 481 PAEDVDTPPRKKRKHRLMAHCRKIQLKDGSSNHYNYQPCDHPQPCDSSCPVIAQ 540  
Db 481 PAEDVDTPPRKKRKHRLMAHCRKIQLKDGSSNHYNYQPCDHPQPCDSSCPVIAQ 540  
Qy 541 NFCEKFCQCSSECCNRFPFGCRCAQCNKTCQPCYLAVERCDPDLCTGGAADHDSKNVS 600  
Db 541 NFCEKFCQCSSECCNRFPFGCRCAQCNKTCQPCYLAVERCDPDLCTGGAADHDSKNVS 600  
Qy 601 CKNCISIOGSKKHLAPSDVAGWGIPIKDPVQKNFISEYCGEIIISQDEADRRGKYVDK 660  
Db 601 CKNCISIOGSKKHLAPSDVAGWGIPIKDPVQKNFISEYCGEIIISQDEADRRGKYVDK 660  
Qy 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKKALQTGE 720  
Db 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKKALQTGE 720  
Qy 721 ELFPDYRSQADALKYVGIEREMEIP 746  
Db 721 ELFPDYRSQADALKYVGIEREMEIP 746

## RESULT 5

US-10-231-778-226

Sequence 226, Application US/10231778

Publication No. US20030126647A1

GENERAL INFORMATION:

APPLICANT: Bildeau, Pierre

APPLICANT: Chaudhury, Abdul M.

APPLICANT: Dennis, Elizabeth S.

APPLICANT: Koltunow, Anna M.G.

APPLICANT: Luo, Ming

APPLICANT: Peacock, William J.

TITLE OF INVENTION: Method for inducing seed development by down-regulating

FILE REFERENCE: 72-98A

CURRENT APPLICATION NUMBER: US/10/231,778

CURRENT FILING DATE: 2002-11-08

PRIOR APPLICATION NUMBER: 09/398,237

PRIOR FILING DATE: 1999-09-20

PRIOR APPLICATION NUMBER: 60/101,184

PRIOR FILING DATE: 1998-09-21

PRIOR APPLICATION NUMBER: AU PP6061

PRIOR FILING DATE: 1998-09-22

PRIOR APPLICATION NUMBER: AU PP6062

PRIOR FILING DATE: 1998-09-22

PRIOR APPLICATION NUMBER: AU PP6063

PRIOR FILING DATE: 1998-09-22

PRIOR APPLICATION NUMBER: AU P01345

PRIOR FILING DATE: 1999-07-01

PRIOR APPLICATION NUMBER: AU P01346

PRIOR FILING DATE: 1999-07-01

NUMBER OF SEQ ID NOS: 239

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 226

LENGTH: 746

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: Ezh1 peptide  
OTHER INFORMATION: fragment  
US-10-231-778-226

Query Match 98.6%; Score 3993; DB 4; Length 746;  
Best Local Similarity 98.3%; Pred. No. 1,5e-279;  
Matches 733; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MGOTGKSEKGPVCMWRKRVSEYVRLQLRFRPADKVFMSNRKILERTETLQNEW 60  
Db 1 MGOTGKSEKGPVCMWRKRVSEYVRLQLRFRPADKVFMSNRKILERTETLQNEW 60  
Qy 61 KORRIQVHLITVSISLRGTRECVTSDDLPTQVLPKLTJANVAVSPIMYSPLQONF 120  
Db 61 KORRIQVHLITVSISLRGTRECVTSDDLPTQVLPKLTJANVAVSPIMYSPLQONF 120  
Qy 121 MVEDETVLANIPIYMGDEVLDQDGTPIBELIKNTDGKYGRCGFIINDEIFVELYNALGQ 180  
Db 121 MVEDETVLANIPIYMGDEVLDQDGTPIBELIKNTDGKYGRCGFIINDEIFVELYNALGQ 180  
Qy 181 YNDDDDDDGDDPDEREREKQLEDRDDEKSRPPRPFPSDKLLEAISMSFPDKGTAEEL 240  
Db 181 YNDDDDDDGDDPDEREREKQLEDRDDEKSRPPRPFPSDKLLEAISMSFPDKGTAEEL 240  
Qy 241 KEKYEKELTEOOLPGALPPECTPNIDGPNKSVOREGSLHSFHTLFCRRCFKYDCFLHPH 300  
Db 241 KEKYEKELTEOOLPGALPPECTPNIDGPNKSVOREGSLHSFHTLFCRRCFKYDCFLHPH 300  
Qy 301 ATPNTYRKKTETALDNKPCGPOCYOHLGAKEPFAALTABRIKTPPRKPGRRGRGLPN 360  
Db 301 ATPNTYRKKTETALDNKPCGPOCYOHLGAKEPFAALTABRIKTPPRKPGRRGRGLPN 360  
Qy 361 NSSRPSPTITNVLSKOTDSREAGTETGSENNDEEEKDETSSESEANSRCQTP1KM 420  
Db 361 NSSRPSPTITNVLSKOTDSREAGTETGSENNDEEEKDETSSESEANSRCQTP1KM 420  
Qy 421 KPNIEPPEVMSGAEASMPFVLIGTYINDFCALRLIGTKCQVYEFKVESSTIAPA 480  
Db 421 KPNIEPPEVMSGAEASMPFVLIGTYINDFCALRLIGTKCQVYEFKVESSTIAPA 480  
Qy 481 PAEDVDTPPRKKRKHRLMAHCRKIQLKDGSSNHYNYQPCDHPQPCDSSCPVIAQ 540  
Db 481 PAEDVDTPPRKKRKHRLMAHCRKIQLKDGSSNHYNYQPCDHPQPCDSSCPVIAQ 540  
Qy 541 NFCEKFCQCSSECCNRFPFGCRCAQCNKTCQPCYLAVERCDPDLCTGGAADHDSKNVS 600  
Db 541 NFCEKFCQCSSECCNRFPFGCRCAQCNKTCQPCYLAVERCDPDLCTGGAADHDSKNVS 600  
Qy 601 CKNCISIOGSKKHLAPSDVAGWGIPIKDPVQKNFISEYCGEIIISQDEADRRGKYVDK 660  
Db 601 CKNCISIOGSKKHLAPSDVAGWGIPIKDPVQKNFISEYCGEIIISQDEADRRGKYVDK 660  
Qy 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKKALQTGE 720  
Db 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKKALQTGE 720  
Qy 721 ELFPDYRSQADALKYVGIEREMEIP 746  
Db 721 ELFPDYRSQADALKYVGIEREMEIP 746

## RESULT 6

US-10-104-047-3162

Sequence 3162, Application US/10104047

Publication No. US20030236392A1

GENERAL INFORMATION:

APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: No. US20030236392A1el full length cDNA

FILE REFERENCE: HI-A0105

CURRENT APPLICATION NUMBER: US/10/104,047

CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3162  
LENGTH: 707  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-3162

Query Match 94.1%; Score 3812.5; DB 4; Length 707;  
Best Local Similarity 94.6%; Pred. No. 1.5e-266;  
Matches 706; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

QY 1 MGOTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSSNRQKILERTTEILNQEW 60  
DB 1 MGOTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSSNRQKILERTTEILNQEW 60  
QY 61 KORRIOPVHILTSVSLRGTRCSVTSDDLPTQVYIPLKTNAAVASVPIIMYSMPLOQNF 120  
DB 61 KORRIOPVHILTSVSLRGTR----- 82  
QY 121 MWEDEVTLHNIPYMGDEVLDODGTFFIEELIKNYDGKVGHDREGCFINDEIFVELVNALQ 180  
DB 83 -VEDEVTLHNIPYMGDEVLDODGTFFIEELIKNYDGKVGHDREGCFINDEIFVELVNALQ 141  
QY 181 YNDDDDDDGDDPEEREKQKQLEDRDDKESRPPKPSDKIFALISSMPDGTAEEL 240  
DB 142 YNDDDDDDGDDPEEREKQKQLEDRDDKESRPPKPSDKIFALISSMPDGTAEEL 201  
QY 241 KEKYELTEOOLPGALPEECTPNIDGPNKSVQREOSLSHFTLFCRCFKYDCFLHPPH 300  
DB 202 KEKYELTEOOLPGALPEECTPNIDGPNKSVQREOSLSHFTLFCRCFKYDCFLHPPH 261  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPRKPGRRGRRLPN 360  
DB 269 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPRKPGRRGRRLPN 321  
QY 361 NSSRSTPTINVLKSDTSDREAGTETGGENNDKEBEKQDETSSSEANSRCQTPIM 420  
DB 322 NSSRSTPTINVLKSDTSDREAGTETGGENNDKEBEKQDETSSSEANSRCQTPIM 381  
QY 421 KENIBPENVEWSGAEAMFRVLIGTYDNFCAIARLIGTKICROYEERVYESSIIAPA 480  
DB 382 KENIBPENVEWSGAEAMFRVLIGTYDNFCAIARLIGTKICROYEERVYESSIIAPA 441  
QY 481 PAEDVDTPPRKKKGRHRLMAAHCRTKIQKDGSSNHVYNYQPCDHPROPCCSSCPCLVIAQ 540  
DB 442 PAEDVDTPPRKKKGRHRLMAAHCRTKIQKDGSSNHVYNYQPCDHPROPCCSSCPCLVIAQ 501  
QY 541 NFCEKPCQCCSSCCQNRPFPCRCCKAQCNTQCFCYLAVERCDPDLCTGCAADHMSKNVS 600  
DB 502 NFCEKPCQCCSSCCQNRPFPCRCCKAQCNTQCFCYLAVERCDPDLCTGCAADHMSKNVS 561  
QY 601 CKNCSIORGSKKHLLIAPSDVAGNGIFITDPOVKKEFTISEYCGEIIISQEARREKRVYDK 660  
DB 562 CKNCSIORGSKKHLLIAPSDVAGNGIFITDPOVKKEFTISEYCGEIIISQEARREKRVYDK 621  
QY 661 YMCSEFLFNLANDFVVDATKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720  
DB 622 YMCSEFLFNLANDFVVDATKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 681  
QY 721 ELFPDYRYSQADALKYVGIEREMEIP 746  
DB 682 ELFPDYRYSQADALKYVGIEREMEIP 707

RESULT 7  
US-11-072-512-3162  
Sequence 3162, Application US/11072512  
Publication No. US20060029945A1  
GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHICO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOYUKI  
APPLICANT: NAGAHARA, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: Novel full length cDNA  
FILE REFERENCE: 08435-0191  
CURRENT APPLICATION NUMBER: US/11/072,512  
CURRENT FILING DATE: 2005-03-07  
PRIOR APPLICATION NUMBER: US 60/350,978  
PRIOR FILING DATE: 2002-01-25  
PRIOR APPLICATION NUMBER: JP 2001-379298  
PRIOR FILING DATE: 2001-11-05  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3162  
LENGTH: 707  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-072-512-3162

Query Match 94.1%; Score 3812.5; DB 6; Length 707;  
Best Local Similarity 94.6%; Pred. No. 1.5e-266;  
Matches 706; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

QY 1 MGOTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSSNRQKILERTTEILNQEW 60  
DB 1 MGOTGKSEKGPVCMRKVKSEYMRRLQKFRRADEVKSMFSSNRQKILERTTEILNQEW 60  
QY 61 KORRIOPVHILTSVSLRGTRCSVTSDDLPTQVYIPLKTNAAVASVPIIMYSMPLOQNF 120  
DB 61 KORRIOPVHILTSVSLRGTR----- 82  
QY 121 MWEDEVTLHNIPYMGDEVLDODGTFFIEELIKNYDGKVGHDREGCFINDEIFVELVNALQ 180  
DB 83 -VEDEVTLHNIPYMGDEVLDODGTFFIEELIKNYDGKVGHDREGCFINDEIFVELVNALQ 141  
QY 181 YNDDDDDDGDDPEEREKQKQLEDRDDKESRPPKPSDKIFALISSMPDGTAEEL 240  
DB 142 YNDDDDDDGDDPEEREKQKQLEDRDDKESRPPKPSDKIFALISSMPDGTAEEL 201  
QY 241 KEKYELTEOOLPGALPEECTPNIDGPNKSVQREOSLSHFTLFCRCFKYDCFLHPPH 300  
DB 202 KEKYELTEOOLPGALPEECTPNIDGPNKSVQREOSLSHFTLFCRCFKYDCFLHPPH 261  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPRKPGRRGRRLPN 360  
DB 262 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPRKPGRRGRRLPN 321  
QY 361 NSSRSTPTINVLKSDTSDREAGTETGGENNDKEBEKQDETSSSEANSRCQTPIM 420  
DB 322 NSSRSTPTINVLKSDTSDREAGTETGGENNDKEBEKQDETSSSEANSRCQTPIM 381  
QY 421 KENIBPENVEWSGAEAMFRVLIGTYDNFCAIARLIGTKICROYEERVYESSIIAPA 480  
DB 382 KENIBPENVEWSGAEAMFRVLIGTYDNFCAIARLIGTKICROYEERVYESSIIAPA 441  
QY 481 PAEDVDTPPRKKKGRHRLMAAHCRTKIQKDGSSNHVYNYQPCDHPROPCCSSCPCLVIAQ 540  
DB 442 PAEDVDTPPRKKKGRHRLMAAHCRTKIQKDGSSNHVYNYQPCDHPROPCCSSCPCLVIAQ 501

```
QY 541 NFCEKFCQCSSECCNRPFGCRCKAQCNTRKQCPCYLAVRECDPDLCTCGAHDMSKNVS 600
|
|
|
Db 502 NFCEKFCQCSSECCNRPFGCRCKAQCNTRKQCPCYLAVRECDPDLCTCGAHDMSKNVS 561
|
|
|
QY 601 CKNCISQKSGSKHLLAPSDVAGWGIFFIKDPVQKNEFFISEYCGEIIISQDEADRGRKYYDK 660
|
|
|
Db 562 CKNCISQKSGSKHLLAPSDVAGWGIFFIKDPVQKNEFFISEYCGEIIISQDEADRGRKYYDK 621
|
|
|
QY 661 YMCSEFLFNANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 720
|
|
|
Db 622 YMCSEFLFNANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGE 681
|
|
|
QY 721 ELFPDYRYSQADALKVYGIEREWEIP 746
|
|
|
Db 682 ELFPDYRYSQADALKVYGIEREWEIP 707
|
|
|
RESULT 8
US-10-153-668-354
; Sequence 354, Application US/10153668
; Publication No. US20030092616a1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STATE Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 354
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-354

Query Match 70.2%; Score 2845; DB 4; Length 517;
Best Local Similarity 100.0%; Pred. No. 8e-197;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 MFPDKGTABELKEKXKLTQQLPGALPPECTPNIDGENAKSVOREGSHSFTLFCRC 289
|
|
|
Db 1 MFPDKGTABELKEKXKLTQQLPGALPPECTPNIDGENAKSVOREGSHSFTLFCRC 60
|
|
|
QY 290 FKYKCFLEHFAATPTTYRKRTETALDNKPCGPOCYOHLSEAKAPPAALTLERIKTPPKR 349
|
|
|
Db 61 FKYKCFLEHFAATPTTYRKRTETALDNKPCGPOCYOHLSEAKAPPAALTLERIKTPPKR 120
|
|
|
QY 350 PGGRRRGRLPNNSSRPSTPTINVLSEKOTDSREAGTETGGENNDKEEEKKDETSSSE 409
|
|
|
Db 121 PGGRRRGRLPNNSSRPSTPTINVLSEKOTDSREAGTETGGENNDKEEEKKDETSSSE 180
|
|
|
QY 410 ANSRQCTPIKMKPNIBPENYEWGASAMFRVLIIGYYDNFCAIARLLIGTKCQYVEF 469
|
|
|
Db 181 ANSRQCTPIKMKPNIBPENYEWGASAMFRVLIIGYYDNFCAIARLLIGTKCQYVEF 240
|
|
|
QY 470 RVKSSSIAPAPAEVDVTPPKKKKKHLLAAHCKKILQKDGSSNNHYNYQPCDHPQP 529
|
|
|
```

```
Db 241 RVKSSSIAPAPAEVDVTPPKKKKKHLLAAHCKKILQKDGSSNNHYNYQPCDHPQP 300
|
|
|
QY 530 CDSSCPVIAQNFCEKFCQCSSECCNRPFGCRCKAQCNTRKQCPCYLAVRECDPDLCTCG 589
|
|
|
Db 301 CDSSCPVIAQNFCEKFCQCSSECCNRPFGCRCKAQCNTRKQCPCYLAVRECDPDLCTCG 360
|
|
|
QY 590 AADHMSKNVSKCNKCS1QRSGSKHLLAPSDVAGWGIFFIKDPVQKNEFFISEYCGEIIISQ 649
|
|
|
Db 361 AADHMSKNVSKCNKCS1QRSGSKHLLAPSDVAGWGIFFIKDPVQKNEFFISEYCGEIIISQ 420
|
|
|
QY 650 EADRGRKYDKYKCSFLFNANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIG 709
|
|
|
Db 421 EADRGRKYDKYKCSFLFNANDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIG 480
|
|
|
QY 710 IPAKRAIQTGEELFPDYRYSQADALKVYGIEREWEIP 746
|
|
|
Db 481 IPAKRAIQTGEELFPDYRYSQADALKVYGIEREWEIP 517
|
|
|
RESULT 9
US-10-115-482-56
; Sequence 56, Application US/10115482
; Publication No. US20030212257A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS
; TITLE OF INVENTION: OF USING THE SAME
; FILE REFERENCE: 21404-322D
; CURRENT APPLICATION NUMBER: US/10/115,482
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/285,325
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: 60/285,890
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286,068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286,292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/287,213
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: 60/288,257
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,134
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/291,725
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/294,771
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/296,965
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/299,128
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 149
; SEQ ID NO 56
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-115-482-56
```

Query Match 60.0%; Score 2429.5; DB 4; Length 723;  
 Best Local Similarity 61.5%; Pred. No. 136-166;  
 Matches 466; Conservative 93; Mismatches 122; Indels 77; Gaps 11;

```

QY 15 WKKRVSEYMLRQLKFRPRADVKSMFSSNRQKILERTILNOEWKORRIPVHILTSV 74
DB 15 WKKRVSEYMLRQLKFRPRADVKSMFSSNRQKILERTILNOEWKORRIPVHILTSV 74
QY 75 SSLRTRCSVTSDL-DPPTQVIPLKTLNVAASVPIMYSWSPLOQNFWEDEVTLNIPY 133
DB 75 SGLPFLKCTISIFPGFASQHMRLMTALVTALVPIWYSWSPLOQNFWEDEVTLNIPY 134
QY 134 MDEVLDDQGTIFELIKYDQKVGNDRE----CGFINDITVELVNALQVNDND---- 185
DB 135 MDEVEKEBDETFIEELINNYDKVHGEEMIPGSLVLSAVLELVDALNOYSDDEEGB 194
QY 186 -----DDDDGDDDEEEREKOKOLEHDDKESRPPKPSDKIFEAISMPDKGTA 237
DB 195 NDTSGKQDDSKEDLPYTRKKRKHAIKGNK--KSKK--KQFPRDITFSALASMPFENGVP 250
QY 238 EELKERYKELTEQOLPGALPPECTPNIDGPNKASVQREGLSHSFHTLFCRCFCYDCFLH 297
DB 251 DDMKERYRELTEMSPBNALPQCTPNIDGPNKASVQREGLSHSFHTLFCRCFCYDCFLH 310
QY 298 PHATPNYKRCOTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPKRGGRRGR 357
DB 311 PHATPNYKRCOTETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPKRGGRRGR 364
QY 358 LPNNSRSPSTPTIN-VLESKQDTSDEAGETETGGENNDKEBEKKOETSSSEANSRCOT 416
DB 365 HHIVASCSNASASAVAEKEDSDRDTG-----NDMASSSEANSRCOT 409
QY 417 PLMKENIETP-----ENVEWSGABASMEFVLIQTYDNFCAIARLIGTKTCROY 467
DB 410 PTKQASAPAPQOLCVHAPSEPEVMTGABESLFRVHGTYFNNFCSIALRLGKTKCKQY 469
QY 468 ERYVRESSIAPAPADVUTPRKKRRLMAAHCRIQLKQDSSNHVINYQPCDHR 527
DB 470 QPAVE-SLILKLPDELMYPSQKKRRLMAAHCRIQLKQDSSNHVINYQPCDHR 528
QY 528 QPCDSSCPVIAQNFCEKQCCSSQCNFPQGCRCQAQNTKQPCYLAVERCDPLCT 587
DB 529 RCDSTCPIMQNFCEKQCCQNP-----LRCDPDLCT 564
QY 588 CGAABHMSKNVSCNKCISIQSGSKHLAPSDVAGWGIPIQPVQKNEFISEYGBEIS 647
DB 555 CGASHMDCKVSCNKCISIQSGSKHLAPSDVAGWGIPIQPVQKNEFISEYGBEIS 624
QY 648 QDEADRGKVDKYNCSFLFNINLPVVDATRKGNKIRFANHSVNPNCYAKVMVNGHR 707
DB 625 QDEADRGKVDKYNCSFLFNINLPVVDATRKGNKIRFANHSVNPNCYAKVMVNGHR 684
QY 708 IGIKRALQGEELPFQVRSOALAKVGIEREHEI 745
DB 665 IGIKRALQGEELPFQVRSOALAKVGIERTDV 722

```

RESULT 10  
 US-11-097-143-14073  
 ; Sequence 14073, Application US/11097143  
 ; Publication No. US20050208558A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Venter, J. Craig  
 ; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
 ; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
 ; FILE REFERENCE: CL000728  
 ; CURRENT APPLICATION NUMBER: US/11/097,143  
 ; CURRENT FILING DATE: 2005-04-04  
 ; PRIOR APPLICATION NUMBER: 60/157,832  
 ; PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: 60/160,191  
 PRIOR FILING DATE: 1999-10-19  
 PRIOR APPLICATION NUMBER: 60/161,932  
 PRIOR FILING DATE: 1999-10-28  
 PRIOR APPLICATION NUMBER: 60/164,769  
 PRIOR FILING DATE: 1999-11-12  
 PRIOR APPLICATION NUMBER: 60/173,383  
 PRIOR FILING DATE: 1999-12-28  
 PRIOR APPLICATION NUMBER: 60/175,693  
 PRIOR FILING DATE: 2000-01-12  
 PRIOR APPLICATION NUMBER: 60/184,831  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: 60/191,637  
 PRIOR FILING DATE: 2000-03-23  
 NUMBER OF SEQ ID NOS: 43008  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 14073  
 LENGTH: 760  
 TYPE: PRT  
 ORGANISM: DROSOPHILA  
 US-11-097-143-14073

Query Match 53.2%; Score 2156.5; DB 6; Length 760;  
 Best Local Similarity 53.9%; Pred. No. 746-147;  
 Matches 432; Conservative 86; Mismatches 176; Indels 107; Gaps 14;

```

QY 8 SEKGVCRKRYKSEYMLRQLKFRPRADVKSMFSSNRQKILERTIL--NOEWKORR 64
DB 3 STKVPEKRRVKSYYIKIRQKRYKRADEIKEMIRWDEHNHVDOLYCSKQWQAKP 62
QY 65 IQPVHILTSVSLRTRCSVTSDDLDPPT--QVPLKTLNVAASVPIMYSWSPLOQNFV 122
DB 63 YDPPHV-----DCVKAQVTSINGISGQPKVPICVINATPLPTMTAPLTOQNFV 115
QY 123 EDETVLANIPYMGDEVLDQGTIFIELIKYDQKVGNDRECGFINDITVELVNAL---- 178
DB 116 EDETVLANIPYMGDEVLDQGTIFIELIKYDQKVGNDRECGFINDITVELVNAL---- 175
QY 179 -----GOYND--DDDDGDDP-----EERE 198
DB 176 SKELERAPGTATATKTETLAKSKQGBDDGVVDADDESPPKLEKTDKGLDTEVEKKE 235
QY 199 KOKLEHRRD-----KESRPRKPSDKITFEATISMPDPDGTAEALBKRYKELTEQOLP 253
DB 226 TEEPLETADAVKPDVSEVKQLPFPAPITFOALSANFPDQTAOLKEKTIETSHQD 295
QY 254 GALPPECTPNIDGPNKASVQREGLSHSFHTLFCRCFCYDCFLHP--HATPNYKRYNT 311
DB 226 -ERPECTPNIDGIAESVSEKRTMHSFHTLFCRCFCYDCFLHRLQGHAGNLOKRYP 354
QY 312 ETALDNKPCGPOCYOHLGAKAFPAALTAERIKTPPKRGGRRGRRLPNNSRSPSTPTIN 371
DB 355 ELKPPAPBCNSNCYVLLIDMEKELAADS--KTPP----- 386
QY 372 VLESKQDTSDEAGETETGGENNDK-----EEREKQETSSSEANSRRCOTPIKQNI 424
DB 387 -----IDSCNEBASSEDSNDSQSNQDPMHENSQKGLTVNSAAVABINSIMAGMNI 440
QY 425 EPEVNEWSGABASMEFVLIQTYDNFCAIARLIGTKTCROYERVRVRESSIAPAPAD 484
DB 441 TSTQCV-WTGADQALYRLVHLVLYLKNYCAIAHMLTKTCROYERVAQGEDAFSEEDRQ 499
QY 485 VDTPEKRRKRLMAAHCRIQLKQDSSNHVINYQPCDHRQPCDSSCPVIAQNFCE 544
DB 500 DFTPEKRRKQRLMSLHCRIQLKQDSSNHVINYQPCDHRGHCDCWNCSCIQOTNFCB 559
QY 545 KFCQSSSCNRPFQGCRCQAQNTKQPCYLAVERCDPLCTLGAABHMSKNVSCNKC 604
DB 560 KFCNCSQCNRPFCRCRCQAQNTKQPCYLAVERCDPLCTQAGC-ADQFLTKTKCNV 618
QY 605 SIQSGSKHLAPSDVAGWGIPIQPVQKNEFISEYGBEISQDEADRGKVDKYNCS 664
DB 619 CVQKGLHGLLMAPSDIAGWGIPLKSGQKNEFISEYGBEISQDEADRGKVDKYNCS 678

```



```

QY      8  SEKGVCRRKVKSEYMRRLQKRRPRADDEVKSMSSNRQKILERTTEIL---NOEMKOR 64
DB      3  STKTPPEKRRKVKSEYIKIRQOKRKRADEIKEMIRWDEHNHNVQULYCESKVMQAKP 62
QY      65  IQPVHILTSVSLRGRECVTSIDLFPPT--QVLEPLKTINAAVAPIMYSWSPLOQNF 122
DB      63  YDPHIV-----DCVKRAEYTSYNGIPSGPOKVPICGINAVTPIPTMYTNAFPOQNFV 115
QY      123  EDEYVLANIPYMGDEVLDQDGTFFIELIKNYDGKVGHDECGFINDIEFVELVNAL--- 178
DB      116  EDEYVLANIPYMGDEVLDQDGTFFIELIKNYDGKVGHDECGFINDIEFVELVNALMSY 175
QY      179  -----GQYND-----DDDDGDDE--BREEKQLEDD----- 205
DB      176  SKLEBAPSTSTAIKTEPLAKSKQGEDGVVDVADACESPMKLEKTESKQDLTDVEKKE 235
QY      206  -----HRDKESRPRK-----PPSDKIFEALISSMFPDKGTAEELKEKYLETEQOLP 253
DB      236  TEETPTEADAVKPAVEEVOKLPPAPLIFQALISANFPDKGTAEELKEKYLETEHQDP 295
QY      254  GALTPECTPNIDGNPAKSVQREOSLHSFHTLFCRCFCYDCTLPF--HATPNTYKROT 311
DB      296  -ERPOECTPNIDGIAESVSRRTMSFHTLPCRCFCYDCTLPF--HATPNTYKROT 354
QY      312  ETALDNKPCGPOCYOHLGAKRPAALTAERIKTTPPKRPGRRRGRRLFNNSRPETPTIN 371
DB      355  ELKPPAEPCNSCYMLDGMKEKLAADS---KTPP----- 386
QY      372  VLESKDTSDEAGTETGGENNDK-----EEREKQDETSSSEANSRCOTPIMKENI 424
DB      387  -----IDSCNEASESDNSNSQPSNKDPFNHNSKQDLTVNSAAVAELNISIMAGMNI 440
QY      425  EPPENVEWSGABSMFVLYIGTYDNFCAIARLIGTKTCROYEYFVRESSIIAPAPAD 484
DB      441  TSJOCV--WTGADQALRYLTHKYVLLKNYCAIAHNMULTKCRQYVERPAQED---AESFSD 496
QY      485  VD---TPRKKRKRHRMAHCRKIQLKQDSSNNVYVYQCDHROQCDSSCPVYON 541
DB      497  LRQDTPPRKKRKRQRLMSLHCRKIQLKQDSSNNVYVYTCDHGHGCDNMCSCTIQON 556
QY      542  FCEKFCQCSSECONRFPGRCCKAQOQNTQCCYLAVERCDPDLCLTCAADHMSKNVC 601
DB      557  FCEKFCQCSSECONRFPGRCCKAQOQNTQCCYLAVERCDPDLCLQACG--AQFPLTKTTC 615
QY      602  KNCISIQSGSKHLLIAPSDVAGWGIPIQDVQKNEFISEYGEIISQDEADRRGVNDY 661
DB      616  KNCVQVGRGLHGLLMAFSDIAGWGIPIKGAQKNEFISEYGEIISQDEADRRGVNDY 675
QY      662  MCSFLFNNNDPVYDATKGNKIRPANSVNPNCYAKMMNNGDHRIGIIPAKRAIQTEB 721
DB      676  MCSFLFNNNDPVYDATKGNKIRPANSINPNCYAKMMVYGDHIGIIPAKRAIQTEB 735
QY      722  LFPDYRYSOADALKVYGIEREEMI 745
DB      736  LFPDYRYGPTQQLKVFGIEREEMI 759

RESULT 13
US-10-424-599-219717
; Sequence 219717, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219717

```

```

; LENGTH: 242
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40432C.1.dep
US-10-424-599-219717

Query Match      29.8%; Score 1208.5; DB 4; Length 242;
Best Local Similarity 91.6%; Pred. No. 5.3e-79;
Matches 230; Conservative 7; Mismatches 5; Indels 9; Gaps 1;

QY      1  MGQTKSEKGPVCMRRKVKSEYMRRLQKRRPRADDEVKSMSSNRQKILERTTEILNDEM 60
DB      1  MGQTKSEKGPVCMRRKVKSEYMRRLQKRRPRADDEVKSMSSNRQKILERTTEILNDEM 60
QY      61  KORRIQPVHILTSVSLRGRECVTSIDLFPPTQVILEPLKTINAAVAPIMYSWSPLOQNF 120
DB      61  KORRIQPVHILTSVSLRGRECVTSIDLFPPTQVILEPLKTINAAVAPIMYSWSPLOQNF 111
QY      121  MVEDEYVLANIPYMGDEVLDQDGTFFIELIKNYDGKVGHDECGFINDIEFVELVNALGO 180
DB      112  MVEDEYVLANIPYMGDEVLDQDGTFFIELIKNYDGKVGHDECGFINDIEFVELVNALGO 171
QY      181  YNDDDDDDGDDPREEKQKQLEHDDKESRPRKFPSPDKIFEALISSMFPDKGTAEEL 240
DB      172  YNDDDDDDGDDPREEKQKQLEHDDKESRPRKFPSPDKIFEALISSMFPDKGTAEEL 231
QY      241  KEKYKELETEQ 251
DB      232  BEKYKELETEQ 242

RESULT 14
US-10-739-930-7656
; Sequence 7656, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739.930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 7656
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C2231_2.p
US-10-739-930-7656

Query Match      20.5%; Score 829; DB 5; Length 895;
Best Local Similarity 29.7%; Pred. No. 7.5e-51;
Matches 240; Conservative 100; Mismatches 217; Indels 252; Gaps 35;

QY      95  VILPKTINAAVAPIMYSWSPLOQNF--FVDEDEYV--HNIPY--MGDEVL--DQDGTFFI 146
DB      142  VVRFVKLPYVERIPPYTTWIFLDKQRMADQSVGRRRIYYDVGNEALICSDSD----- 197
QY      147  BELIKNYDGK---VHGQ-----RCCGFINDIEFVELVNALGOYNDDDDDGDGDPDEE 195
DB      198  BEIPEPEBEKGFPTGSDQLWRATQHRG--LNREV---VNVLCQFID-----STPSE 245
QY      196  REEKQKLEHDDKESRPRKFPSPDKIFEALISSMFPDKGTAEELKEKYLETEQOLPGA 255
DB      246  IEHSVTLFR--KKNKNG-----SSDKI-----ERQ--- 271
QY      256  LPPECTPNIDGNPAKSVQREOSLHSFHTLFCRCFCYDCTLPF--HATPNTYK 307
DB      272  -----SLD---KTM--DAVIDSFNLFGRCLVDCRLHGSQNLVFPTEKQPYSF- 317
QY      308  RKNTETALDNKPCGPOCYOHLGEG----- 330

```

Db 318 -----EPDENKKPCGRCQCYLRWGRGFORIHVGLSGCATYNNMESGTSHKVDVSIIMSESD 373  
; : ||| |||  
Qy 331 -----AKEPALTAERIKTPP-----KR 349  
; : : : :  
Db 374 SNREKGNIRMTLVGTSGSKLISVSABESTTPPSATSTETENASSMPSPSLKRYISK 433  
; : : : :  
Qy 350 PGGRRRGRLPNNSSRPSTPTI---NYL-----ESKDTSDRBAETGTGENNDKEEE 399  
; : : : :  
Db 434 RGRPRRSPGKQKQVFTSDISFASNTLNKLISPEIHD---RLESREPG---DKLQ 465  
; : : : :  
Qy 400 KQDTSSSSSEANSRC-OTPIKKNPI-EPEENVE-----WSGABASFRVLIG 445  
; : : : :  
Db 486 ILDESTKTSKSDICGSESPITTTENVERESKVSSTNNFLEHTLSCWALSERDLYLKIG 545  
; : : : :  
Qy 446 TYDNPFCALR--LIGTKCRQYVERFRVKSIIAPAP-----ADV 485  
; : : : :  
Db 546 IFGNISCLIANLISGMKTCHEVANYMYNNGAMAKRPLNKSISGDPATETEDYMEQDM 605  
; : : : :  
Qy 486 DTEPRKKKKRHR-----LMAHCRKIQLKKDSSSNHY-NYOPCDHPRQPCDSSCPYI 538  
; : : : :  
Db 606 VARRIRYRRGRNKKLKYTWKSAGHPYRKRIQDGKQWYQYNPC-VCOQCGKDCQVE 664  
; : : : :  
Qy 539 AQNFCERFCCSSCCQRPFQCRG-KAQCNTKQCPCTLAVERCDPDLCTCGAADHWS- 596  
; : : : :  
Db 665 NGTCCEKYCGCSKSKNFRGCHCAKQCSRCQCPFAASRECDPVCRCNC-----WVSC 719  
; : : : :  
Qy 597 -----KNVSCNKGSIQSGSKHLLAPSDVAGMGIPIKDPVQKNRPISEYCGE 644  
; : : : :  
Db 720 GDGSLGEPAPAGDYGCGNMKLLKQOORILGSDVAGWCAPIKNFVNNDYIGETGE 779  
; : : : :  
Qy 645 IISQDEADRRGKYVDKTYMCSFLFNIMNDFVVDATRKGNKIRFANHSVNPNCYATVMVNG 704  
; : : : :  
Db 780 LISHKADKRGKTYDRANSSFLFDLNDQYLDAYRKDGKLFANHSNPNCYATVMVAG 839  
; : : : :  
Qy 705 DHRIGIFAKRAIQTEBELFPDYRYSQADA 733  
; : : : :  
Db 840 DHRVGIYAKHEIEASEBELFYDYRYGPDQA 868  
; : : : :  
RESULT 15  
US-10-231-778-221  
; Sequence 221, Application US/10231778  
; Publication No. US20030126647A1  
; GENERAL INFORMATION:  
; APPLICANT: Bildeau, Pierre  
; APPLICANT: Chaudhury, Abdul M.  
; APPLICANT: Dennis, Elizabeth S.  
; APPLICANT: Loo, Ming  
; APPLICANT: Koltunow, Anna M.G.  
; APPLICANT: Peacock, William J.  
; TITLE OF INVENTION: Method for inducing seed development by down-regulating  
; TITLE OF INVENTION: expression of the F182 gene  
; FILE REFERENCE: 72-98A  
; CURRENT APPLICATION NUMBER: US/10/231,778  
; CURRENT FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: 09/398,237  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 60/101,164  
; PRIOR FILING DATE: 1998-09-21  
; PRIOR APPLICATION NUMBER: AU P6061  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: AU P6062  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: AU P6063  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: AU P60345  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: AU P60346  
; PRIOR FILING DATE: 1999-07-01  
; NUMBER OF SEQ ID NOS: 239  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 221

; LENGTH: 856  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: E2A1 peptide  
; OTHER INFORMATION: fragment  
US-10-231-778-221  
Query Match 20.5%; Score 828.5; DB 4; Length 856;  
Best Local Similarity 28.8%; Pred. No. 7.7e-51;  
Matches 251; Conservative 113; Mismatches 267; Indels 241; Gaps 38;  
Qy 25 RLQQLKRRADRYKSM---FSSNRQKILERTILNQMCKRRIQVPHILTSVSLRGTR 81  
; : : : :  
Db 35 RLSELNRKRIQGEKVRISIKKEFENRRKV-----DAHVSF---FSSAASSRATA 79  
; : : : :  
Qy 82 EGSVTSDDLPPTQVPIPLKTLNAAVAVPIMWSPELQONFVWDEVTLANI-----P 132  
; : : : :  
Db 80 EDNGNSMTLSRRKRMPLCTLNG-----FSGHVGDRDYPTKDVISASVYKLPARIRIP 132  
; : : : :  
Qy 133 YMGDEVLDQDGTPIELIKNYDGKVGHDRECGFINDEIFVELYNALQVINDD--DDDG 190  
; : : : :  
Db 133 YTWIFLDRNQMAR-----DQSVVGR-----QIYYE-----QHGGETLIGSDSE 173  
; : : : :  
Qy 191 DDEPEREREKQDLIEDHRDDKESRPRKFPEDKIFEAISSFPDKGTABELK----- 241  
; : : : :  
Db 174 EEPERPEREKRFSEG-----EDSIITWLGQRY---GMGEYQDALCOLLSV 216  
; : : : :  
Qy 242 -----EKYELTEOQLPGLPPECTPNIDPNAK-SVQREQ---SLHSFHTLFCGRCP 290  
; : : : :  
Db 217 DASDILERYNELKLD-----KQNTSEFNSGFKIGLSKLGALDSFDLFCGRCL 270  
; : : : :  
Qy 291 KYDCFLHFPFATNTYRKKN---TETALDNKPCGPOCYOLHBAKE-----PAA--- 336  
; : : : :  
Db 271 VFDCLHGCQPLISASEKQPYMSDYEGDKRCPCKHCYQLQKAVREVPETCSFASVAAE 330  
; : : : :  
Qy 337 -----ALTAER---IKTPPRKPG-----GRRR----- 355  
; : : : :  
Db 331 KASEBECSKAVSDVPHPAASGVSLQVEKTDIGIKVNDSSSGVQEHGIRGKBVPIKLD 390  
; : : : :  
Qy 356 -GRLPNNSRP-----STPTNVLESKQDSDRBAETGTGENNDKEEREK 400  
; : : : :  
Db 391 SNDLPNLSNKKQKTAAADTMSFVNSVPSLD--QALDSTGDDGCTTDNKNVNDSEDAK 448  
; : : : :  
Qy 401 K-----DETSSSEANSRCQTP-----IKMKRNIEPPENVEWSGABASFRVLIG 445  
; : : : :  
Db 449 EVGEPIDNVHOGSSICQPHHSGNGALIIIAEMSETSRP--STEWNPIEKDLYLKIGE 506  
; : : : :  
Qy 446 TYDNPFCALR--LIGTKCRQYVER-RVYESGIAPA-----PAEDVD--TP 488  
; : : : :  
Db 507 IFGNISCLIANLISGLKTLDVSNYRNEVESVFRSSTPNLLDGRTPGNDNDVBP 566  
; : : : :  
Qy 489 PR-----KKKRRHRLM-----AAHCRKIQLKKDSSSNHYNYOPCDHPRQPCDSSCPYI 538  
; : : : :  
Db 567 PRTLPFRKRGKTRKLTAKSTYSAGHPSWKRIAGKNSCKQYTPCG-CLSMCGDCCCLT 625  
; : : : :  
Qy 539 AQNFCERFCCSSCCQRPFQCRG-KAQCNTKQCPCTLAVERCDPDLCTCGAADHWSK 597  
; : : : :  
Db 626 NETCCEKYCGCSKSKNFRGCHCAKQCSRCQCPFAASRECDPVCRCNC-----W--- 677  
; : : : :  
Qy 598 NVSCNKGSI---QRC-----SKHLLAPSDVAGMGIPIKDPVQKNRPISEY 641  
; : : : :  
Db 678 -VSCGDLSLBAPRGGGCGNNMLLIRQOORILGSDVAGWCAPIKNFVNNDYIGETGE 736  
; : : : :  
Qy 642 CGEIIISQDEADRRGKYVDKTYMCSFLFNIMNDFVVDATRKGNKIRFANHSVNPNCYATVM 701  
; : : : :  
Db 737 TGEILISHKADKRGKTYDRANSSFLFDLNDQYLDAYRKDGKLFANHSKPNYATVMF 796  
; : : : :  
Qy 702 VNGDHRIGIFAKRAIQTEBELFPDYRYSQADA 733  
; : : : :  
Db 797 VAGDHRVGIYANERIEASEBELFYDYRYGPDQA 828  
; : : : :



Mon Aug 14 10:47:45 2006

us-10-773-302-2.rapbm

Page 11

Search completed: August 11, 2006, 22:49:53  
Job time : 131.001 secs

---

46\$ PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:45:55 ; Search time 23.2318 Seconds

(without alignments)  
2161.464 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGQTGKSEKGPVCMRKRVK.....RYSQADALKYVGIRERMEIP 746

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 239914 seqs, 67312017 residues

Total number of hits satisfying chosen parameters: 239914

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New\*

1: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*  
2: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*  
3: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*  
4: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*  
5: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*  
6: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US10\_NEM\_PUB.pep.\*  
7: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US10\_NEM\_PUB.pep.\*  
8: /EMC Celerra\_SIDS3/ptodata/2/pubpaa/US60\_NEM\_PUB.pep.\*

\* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4030.5	99.5	751	6	US-10-539-630-1 Sequence 1, Appli
2	828	20.4	777	6	US-10-953-349-35469 Sequence 35469, A
3	828	20.4	895	6	US-10-953-349-35468 Sequence 35468, A
4	828	20.4	913	6	US-10-953-349-35467 Sequence 35467, A
5	809.5	20.0	894	6	US-10-953-349-37881 Sequence 37881, A
6	809.5	20.0	903	6	US-10-953-349-37880 Sequence 37880, A
7	806.5	19.9	874	6	US-10-953-349-37882 Sequence 37882, A
8	769	19.0	895	7	US-11-330-403-9237 Sequence 9237, Ap
9	299.5	7.4	497	7	US-11-330-403-16696 Sequence 16696, A
10	297	7.3	511	6	US-10-953-349-35696 Sequence 35696, A
11	297	7.3	513	6	US-10-953-349-35695 Sequence 35695, A
12	297	7.3	551	6	US-10-953-349-35694 Sequence 35694, A
13	291	7.2	425	7	US-11-330-403-11703 Sequence 11703, A
14	287.5	7.1	369	7	US-11-330-403-9402 Sequence 9402, Ap
15	286.5	7.1	980	7	US-11-330-403-2871 Sequence 2871, Ap
16	279.5	6.9	916	6	US-10-539-228-355 Sequence 355, App
17	276	6.8	437	6	US-11-330-403-13362 Sequence 13362, A
18	276	6.8	612	6	US-10-449-902-47628 Sequence 47628, A
19	275.5	6.8	352	7	US-11-330-403-14838 Sequence 14838, A
20	275.5	6.8	407	7	US-11-330-403-6394 Sequence 6394, Ap
21	273.5	6.8	393	7	US-11-330-403-10441 Sequence 10441, A
22	273.5	6.8	585	7	US-11-330-403-5638 Sequence 5638, Ap
23	273	6.7	954	7	US-11-330-403-11849 Sequence 11849, A
24	272.5	6.7	344	7	US-11-330-403-2509 Sequence 2509, Ap
25	272.5	6.7	363	7	US-11-330-403-13204 Sequence 13204, A

25	272	6.7	1212	6	US-10-449-902-46235 Sequence 46235, A
27	270.5	6.7	1051	7	US-11-330-403-8440 Sequence 8440, Ap
28	270	6.7	362	7	US-11-330-403-6304 Sequence 6304, Ap
29	269.5	6.7	339	7	US-11-330-403-15444 Sequence 15444, A
30	269	6.6	340	6	US-10-449-902-31539 Sequence 31539, A
31	265.5	6.6	477	7	US-11-330-403-14719 Sequence 14719, A
32	265.5	6.6	844	7	US-11-330-403-4780 Sequence 4780, Ap
33	264.5	6.5	870	6	US-10-539-228-358 Sequence 358, App
34	264	6.5	340	7	US-11-330-403-35229 Sequence 5229, Ap
35	264	6.5	363	7	US-11-330-403-18820 Sequence 18820, A
36	260.5	6.4	349	7	US-11-330-403-13539 Sequence 13539, A
37	256.5	6.3	349	7	US-11-330-403-394 Sequence 394, A
38	256	6.3	292	7	US-11-330-403-16378 Sequence 16378, A
39	256	6.3	317	7	US-11-330-403-3143 Sequence 3143, Ap
40	256	6.3	342	7	US-11-330-403-460 Sequence 480, App
41	256	6.3	342	7	US-11-330-403-4107 Sequence 4107, Ap
42	256	6.3	786	7	US-11-330-403-1643 Sequence 1643, Ap
43	255	6.3	412	6	US-10-511-937-2580 Sequence 2580, Ap
44	254	6.3	637	6	US-10-449-902-35907 Sequence 35907, A
45	252.5	6.2	684	7	US-11-330-403-6172 Sequence 6172, Ap

## ALIGNMENTS

RESULT 1					
US-10-539-630-1					
; Sequence 1, Application US/10539630					
; Publication No. US20060104981A1					
; GENERAL INFORMATION:					
; APPLICANT: Takeda Chemical Industries, Ltd.					
; TITLE OF INVENTION: Preventing and treating agent for cancer					
; FILE REFERENCE: 3130WOP					
; CURRENT FILING DATE: 2005-06-17					
; PRIOR APPLICATION NUMBER: JP2002-373144					
; NUMBER OF SEQ ID NOS: 14					
; SEQ ID NO 1					
; LENGTH: 751					
; TYPE: PRT					
; ORGANISM: Human					
US-10-539-630-1					
Query Match					
Best Local Similarity 99.5%; Score 4030.5; DB 6; Length 751;					
Matches 745; Conservativity 0; Pred. No. 2.3e-243;					
Matches 1; Indels 5; Gaps 1;					
QY	1	MGQTGKSEKGPVCMRKRVKSEYMLRLQKPRRADVKSMSSNRKILERTILNDEW	60		
DB	1	MGQTGKSEKGPVCMRKRVKSEYMLRLQKPRRADVKSMSSNRKILERTILNDEW	60		
QY	61	KORRIQPHILTSVSLGTRKCSVTSDLPPTQYIPLKTLNAVASVIMYSWSPLOQNF	120		
DB	61	KORRIQPHILTSVSLGTRKCSVTSDLPPTQYIPLKTLNAVASVIMYSWSPLOQNF	120		
QY	121	MVEDETVLANIPLYMGDEVLDQGTIFBELINNYDGKVGHBRCGFINDEIFELVNALGQ	180		
DB	121	MVEDETVLANIPLYMGDEVLDQGTIFBELINNYDGKVGHBRCGFINDEIFELVNALGQ	180		
QY	181	YNDDDDDDDGDDPEREREKQDLBHRDDKSRPRKPSDKIFPAISMPDPKTAEBL	240		
DB	181	YNDDDDDDDGDDPEREREKQDLBHRDDKSRPRKPSDKIFPAISMPDPKTAEBL	240		
QY	241	KBYKELTEQQLPGALPEPCTPNIDGPNKSVOROSLHSPHTLPCRCFCFKDCFLH---	297		
DB	241	KBYKELTEQQLPGALPEPCTPNIDGPNKSVOROSLHSPHTLPCRCFCFKDCFLH---	297		
QY	298	--PHATPNTYKRRKNTETALDNKCGPQCYOHLBKAKFAAALTAERIKTPPKRGGR	355		
DB	298	--PHATPNTYKRRKNTETALDNKCGPQCYOHLBKAKFAAALTAERIKTPPKRGGR	355		
QY	301	NYSFATPNTYKRRKNTETALDNKCGPQCYOHLBKAKFAAALTAERIKTPPKRGGR	360		
DB	301	NYSFATPNTYKRRKNTETALDNKCGPQCYOHLBKAKFAAALTAERIKTPPKRGGR	360		
QY	356	GRLPNRSRPSPTPTINTVLSKDTDSDRAGTGTGGENNDKEBEKDKDTSSSSSANSQC	415		

```
Db 361 |GRLPNSSRPSPTPTINVLSEKOTSDREAGTETGGENNDKEEKKOETSSSSSEANSRCQ| 420
Qy 416 |TPIMKPIIEBPENVENSAGASMPFVLIGYYNFCALIGTKICROYEPRVYESS| 475
Db 421 |TPIMKPIIEBPENVENSAGASMPFVLIGYYNFCALIGTKICROYEPRVYESS| 480
Qy 476 |IAPAPADVDTPRKKRKRHLMAAHCKRIQLKKDGSNNHYVYQPCDHPRQPCDSSCP| 535
Db 481 |IAPAPADVDTPRKKRKRHLMAAHCKRIQLKKDGSNNHYVYQPCDHPRQPCDSSCP| 540
Qy 536 |CVALQNFCEKQCQSSSECONRPFPGCRCAQCNTRKQPCYLAVRECDPDLCTCGAADM| 595
Db 541 |CVALQNFCEKQCQSSSECONRPFPGCRCAQCNTRKQPCYLAVRECDPDLCTCGAADM| 600
Qy 596 |SKVNSCKNCSIORGSKHLLAPSDVAGMGIPIKDPVQKNFISEYCEIISODEADRG| 655
Db 601 |SKVNSCKNCSIORGSKHLLAPSDVAGMGIPIKDPVQKNFISEYCEIISODEADRG| 660
Qy 656 |KYVDKYMCSFLNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIPAKRA| 715
Db 661 |KYVDKYMCSFLNLNDFVVDATRKGNKIRFANHSVNPNCYAKVMVNGDHRIGIPAKRA| 720
Qy 716 |IQTGBELFPDYRYSQADALKYVGIEREMEIP 746
Db 721 |IQTGBELFPDYRYSQADALKYVGIEREMEIP 751
```

## RESULT 2

```
US-10-953-349-35469
; Sequence 35469, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREX
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35469
; LENGTH: 777
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35469
```

Query Match 20.4%; Score 828; DB 6; Length 777;

Best Local Similarity 29.5%; Pred. No. 6.5e-44;

Matches 240; Conservative 100; Mismatches 213; Indels 260; Gaps 35;

```
Qy 95 |VPLKTLNAVASVPIWMSPLQON-FWVEDETVL--HNIPY--MGDEVL---DQDGTFI 146
Db 24 |VVRVFKLPLVERIPYTTWIFLDKNQMAADQSVGRRRIYYDPVGNELICSDSD--- 79
Qy 147 |ELIKNYDGK---VHGD-----RBCGFINDEIFVELVNALGQYNDDDDDDDGDDPEE 195
Db 80 |EILPEPEBEKHFTEGEGQILWRATQEHG-LNREV---VVNLQCFID-----STPSE 127
Qy 196 |REBKQDLIEDHDDKESRPPKFPSPDKIFEAISSMPFDKGIAELIKETKELTEQOLPGA 255
Db 128 |IEERSEVLFE-KNEKNSG-----SSDKI-----EROL--- 153
Qy 256 |LPPECTPINIDGNASVQRBOSLHSFHTLFCRRCFKYDCFLH-----PFHATPNTYK 307
Db 154 |-----SLD---KTM---DAVLDSFDNLFCRCLVFCRHLHGSQNLVFPTEKQYISF- 199
Qy 308 |RKNTEALDNKFCGPGCYQHLEG----- 330
Db 200 |-----EPDENKKPCGCGCTLRMRGFGQELHDVGLSCATYNNMSGTVSHKVDVSISSSED 255
Qy 331 |-----AKEPAAALTAERIKTP-----KR 349
```

```
Db 256 |SNREKINRSMTLVGTSISKIISVSAEBSSTPPSADTSTENASSMPSSLRKYISK 315
Qy 350 |PGRRRGRRLPNSSRPSPTPTI---NVL-----ESKOTSDREAGTETGGENNDKEE 399
Db 316 |RGRPRRERSPGKQKVTSISFASNTLNTLSIPEIDT---RLESREPGG---DKLQ 367
Qy 400 |KQDETSSSSSEANSRC-OTPIKMKPIEPPENVE-----WSGAEASMR 441
Db 368 |ILDESTKTSKSDICGSSPIYTTENM---GISBKVYSTKNFLEHTLSGWSALERDYL 423
Qy 442 |VLIGTYDNFCATAR--LIGTKICROYEPRVYESSIAPAP----- 481
Db 424 |KGLIEFGNSCLLARNLISGKTCMEVANYMNGAMAKRPLNKSISGDPAETBODYM 483
Qy 482 |AEDVTPPRKKRKR-----LMAAHCKRIQLKKDGSNNHY--NYQPCDHPRQPCDSSC 534
Db 484 |EQDMVARTRIYRRGRNRKLYTKWSAGHPTVRKRIGDKQWQTYNPF-VCCQMGCKDC 542
Qy 535 |PCVALQNFCEKQCQSSSECONRPFPGCRCAQCNTRKQPCYLAVRECDPDLCTCGAADH 593
Db 543 |PCVENGTCCCKYCGCSXSKNKRFGCHCAQCSQRCPCFASRECDPVCENC----- 597
Qy 594 |WDS-----KNVSCNCSIORGSKHLLAPSDVAGMGIPIKDPVQKNFISE 640
Db 598 |WVSCGDSLSGPPRARGCYQCGNWKLLKQOQRIILRSVYAGMGATIKFVANKNDYLGE 657
Qy 641 |YCGEIIISODEADRGKRYDKYMSFLNLNDFVVDATRKGNKIRFANHSVNPNCYAKVM 700
Db 658 |YTGBELIHKEDAKKGIYDRANSFLFDLNDQYLDAYRKGDKLKFANHSNPNCYAKVM 717
Qy 701 |MWDGDRHIGIFAKRAIQTGBELFPDYRYSQADA 733
Db 718 |LVAGDHRVGIYAKKHIEASELFPDYRYPDQA 750
```

## RESULT 3

```
US-10-953-349-35468
; Sequence 35468, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREX
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35468
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35468
```

Query Match 20.4%; Score 828; DB 6; Length 895;

Best Local Similarity 29.5%; Pred. No. 7.6e-44;

Matches 240; Conservative 100; Mismatches 213; Indels 260; Gaps 35;

```
Qy 95 |VPLKTLNAVASVPIWMSPLQON-FWVEDETVL--HNIPY--MGDEVL---DQDGTFI 146
Db 142 |VVRVFKLPLVERIPYTTWIFLDKNQMAADQSVGRRRIYYDPVGNELICSDSD--- 197
Qy 147 |ELIKNYDGK---VHGD-----RBCGFINDEIFVELVNALGQYNDDDDDDDGDDPEE 195
Db 198 |EILPEPEBEKHFTEGEGQILWRATQEHG-LNREV---VVNLQCFID-----STPSE 245
Qy 196 |REBKQDLIEDHDDKESRPPKFPSPDKIFEAISSMPFDKGIAELIKETKELTEQOLPGA 255
Db 246 |IEERSEVLFE-KNEKNSG-----SSDKI-----EROL--- 271
Qy 256 |LPPECTPINIDGNASVQRBOSLHSFHTLFCRRCFKYDCFLH-----PFHATPNTYK 307
Db 272 |-----SLD---KTM---DAVLDSFDNLFCRCLVFCRHLHGSQNLVFPTEKQYISF- 317
```

```

QY 308 RNTETALDNKPGQCYOHLG----- 330
Db 318 ----EPDENKKPCGRCQCYLRMRGQFOEIHVGLSGCATYNNESGTVSHKVDVINSSESD 373
QY 331 -----AKFAALTAARIKTP-----KR 349
Db 374 SNREKNIIRSMTLVGTSSKTIISVASESTTTPPADTSETNASSDMPSSLKRYKISK 433
QY 350 PGRRRRGLPNNSSPSTPTI---NVL-----ESKOTDSDEAGTETGENNDKEEBE 399
Db 434 RQPRYRERSPGKQKQVFTSDISFASNIINKLSIPEIRDT---RLSREBEG---DKLQ 485
QY 400 KQDETSSSSSEANSRC-QTPIKMKPIPEPNE-----WGASAMFR 441
Db 486 ILDESTKKTSSKDTCGESPITTTENM---GIESKKVSTKNFLEHTLSCWALERDYL 541
QY 442 VLIQTYDNFCAIAR--LIGTKCROYEFYVESIILAPAP----- 481
Db 542 KQIEIFGKNSCLIAIRNLGSKTICMEVANYMNGAAMAKRPLANKSISGDAFETBODYM 601
QY 482 AEDVTPPKKKRKR-----LMAAHCRIQLKDGSSNHVY-NYQPCDHPROPDSSC 534
Db 602 BQDMVARTIRYRRGRNRKLKTYTKSAGHPYRKIRIGDKQWYQYNPC-VCCQMGKQK 660
QY 535 PCVIAQNFCEKFCQSSBECQNFPGRC-KAQQNTQCPCYLAVERCDLCLTGAAAH 593
Db 661 PCVENGTCCCKYCGSKSCKNFRGCHCAKSCQRCPCFASRECDPDVCRC----- 715
QY 594 WDS-----KNVSCKNCSIORGSKKHLIAPSDVAGWGIPIQPKNEFISE 640
Db 716 WISCQDGLGEPAPADGQCGNMILLKQQRILIGRSDVAGWGAFINPNKNDYLG 775
QY 641 YGGEIISQDEADRGKVDYKTCMCSFLFNINDFVVDATRKGNKIRFANHSVNPNCYAKYM 700
Db 776 YTGELISHKADRGKRIYDRANSSFLFDINDQYLDAYARKGDKLKFANHSNPNCAKYM 835
QY 701 MNGDHRIGIFAKRAIQGEBELFPDRIYSGADA 733
Db 836 ILAGDHRVGIIYAKHEIASEBELFYDYRYGPDQA 868

RESULT 4
US-10-953-349-35467
; Sequence 35467, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDS
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35467
; LENGTH: 913
; TYPE: PR1
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35467

Query Match 20.4%; Score 828; DB 6; Length 913;
Best Local Similarity 29.5%; Pred. No. 7.8e-44;
Matches 240; Conservative 100; Mismatches 213; Indels 260; Gaps 35;

```

```

QY 196 REEKQJEDHRDDKESRPPKRPEDKIFEAISWFPDKTAELKKEYKELTBOQLPGA 255
Db 264 IEEERSEVFE-KNENSG-----SSDKI-----ERQL----- 289
QY 256 LPPECTPNIIDGNASVORBOGLSHFHTLFCRCRCRYOCFLH-----PHAFNPNTYK 307
Db 290 -----SLD-----KTM-DAVLSFDNLFCRCRLVFDORLHGCSQNLVFPTEKQPYSP- 335
QY 308 RNTETALDNKPGQCYOHLG----- 330
Db 336 ----EPDENKKPCGRCQCYLRMRGQFOEIHVGLSGCATYNNESGTVSHKVDVINSSESD 391
QY 331 -----AKFAALTAARIKTP-----KR 349
Db 392 SNREKNIIRSMTLVGTSSKTIISVASESTTTPPADTSETNASSDMPSSLKRYKISK 451
QY 350 PGRRRRGLPNNSSPSTPTI---NVL-----ESKOTDSDEAGTETGENNDKEEBE 399
Db 452 RQPRYRERSPGKQKQVFTSDISFASNIINKLSIPEIRDT---RLSREBEG---DKLQ 503
QY 400 KQDETSSSSSEANSRC-QTPIKMKPIPEPNE-----WGASAMFR 441
Db 504 ILDESTKKTSSKDTCGESPITTTENM---GIESKKVSTKNFLEHTLSCWALERDYL 559
QY 442 VLIQTYDNFCAIAR--LIGTKCROYEFYVESIILAPAP----- 481
Db 560 KQIEIFGKNSCLIAIRNLGSKTICMEVANYMNGAAMAKRPLANKSISGDAFETBODYM 619
QY 482 AEDVTPPKKKRKR-----LMAAHCRIQLKDGSSNHVY-NYQPCDHPROPDSSC 534
Db 620 BQDMVARTIRYRRGRNRKLKTYTKSAGHPYRKIRIGDKQWYQYNPC-VCCQMGKQK 678
QY 535 PCVIAQNFCEKFCQSSBECQNFPGRC-KAQQNTQCPCYLAVERCDLCLTGAAAH 593
Db 679 PCVENGTCCCKYCGSKSCKNFRGCHCAKSCQRCPCFASRECDPDVCRC----- 733
QY 594 WDS-----KNVSCKNCSIORGSKKHLIAPSDVAGWGIPIQPKNEFISE 640
Db 734 WISCQDGLGEPAPADGQCGNMILLKQQRILIGRSDVAGWGAFINPNKNDYLG 793
QY 641 YGGEIISQDEADRGKVDYKTCMCSFLFNINDFVVDATRKGNKIRFANHSVNPNCYAKYM 700
Db 794 YTGELISHKADRGKRIYDRANSSFLFDINDQYLDAYARKGDKLKFANHSNPNCAKYM 853
QY 701 MNGDHRIGIFAKRAIQGEBELFPDRIYSGADA 733
Db 854 ILAGDHRVGIIYAKHEIASEBELFYDYRYGPDQA 886

RESULT 5
US-10-953-349-37881
; Sequence 37881, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDS
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37881
; LENGTH: 894
; TYPE: PR1
; ORGANISM: Zea mays subsp. mays
US-10-953-349-37881

Query Match 20.0%; Score 809.5; DB 6; Length 894;
Best Local Similarity 26.8%; Pred. No. 1.1e-42;
Matches 260; Conservative 108; Mismatches 249; Indels 353; Gaps 42;
QY 3 QTGKSKSG-----PVCNRKRVKSEYRMLRQLKFRPADVKSMPFSNRQKILE 51

```



```

QY 604 CSIQRSKXHLILAPSDVAGMGIFIKDPVQKNEFISEYCGEILISODEADRGKVYKXMC 663
DB 747 MKLLKQOQRIILGRSDVAGMGAFIKPNVKNNDYLGEYIGELISHREADKRGKITVRANS 806
QY 664 SPFLNANDPVVDATRGKNIKIRFANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGEELF 723
DB 807 SPFLNDNDQVYLDAYRKGDKIKFANHSNPNCYAKVMVAGDHRVGIYAKHEILASEELF 866
QY 724 PDYRYSQADA 733
DB 867 YDYRYPDQA 876

RESULT 7
US-10-953-349-37882
; Sequence 37882, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37882
; LENGTH: 874
; TYPE: PRF
; ORGANISM: Zea mays subsp. mays
US-10-953-349-37882

Query Match
Best Local Similarity 19.9%; Score 806.5; DB 6; Length 874;
Matches 252; Conservative 103; Mismatches 244; Indels 325; Gaps 39;

QY 38 VASMSBSNRKILERTIELNQEKKQR---IQPVHILTSVSLKGT-----RE 82
DB 21 IQVOSGLAYIKKLEKLV-NRKTLOHSGSLPDVAAAASVAS-RGTDGGNALSORAAEQ 78
QY 83 CSVTSDL-----DPTVOVPLKTLNVA-----SVPLMWSM 113
DB 79 CG-SDLANGIGERDVSVQBERLATGTLALSSGATARTYRPAKPLVEKIPYTTW 136
QY 114 SPLQON-FWVEDETVL--HNIPY--MGDEVL--DODGTFTB-----LINKYDGK-- 157
DB 137 IFLDKNQRMADQSVGRIRIYDVTGNALICSDSDEIIPREBEKHFTGGBDLIMR 196
QY 158 ---HGRECGFINDERIFVELYNALGOYNDDDDDGDDPPEREREKQKULEHRRDKESR 213
DB 197 ATQDHH-----LNQEV-----VNVLCQF-----IGATPSIEERSEVLFK-KNEKHS 238
QY 214 PRKFPSS---DKIFEALISWFPDGTAELEKRYELTEQQLPGALPPECPTNIDGPVA 269
DB 239 SSDKIESRLSLDKTMDAV----- 256
QY 270 KSVQBEQSLHSFHTLFCRCFCFYKDCFLH-----PFAATPNTYKRNKTETALD 316
DB 257 ---LDSFDNLFCRCCLVDFCDLHGCSQNLVFPCEKQYSPDPBNK----- 299
QY 317 NKPCGPOCY---OHLBKAKE----- 333
DB 300 -FPCHLCTLRPPQWRBEGFKEMHDDGLAGATYTWESGASQRVUVNVMYSESDSNRQK 358
QY 334 -----FAAATLARIKTPP-----KRPGRRR 355
DB 359 NTRSMVLVGTSGSKIISSVASESTTTPADISETENVSDDLPPSLRKHAKISKGPFRK 418
QY 356 GRLPNNSRRSTPTI-----NVLESKOTDSREAGTETGGENNDKEBEKKDETS 405
DB 419 BHSFGKRGVFTSDISFEGNIMNKLISPIEIRDT---RLBSRESG-----DKRLIDSET 470

```

```

QY 406 SSSBANSRC-OTPIKKKPIEPPEN-----VE-----WSGABASMFRLIGTYNDP 451
DB 471 KTSRKDMCGSPATTMENVGRQSNKYVSTKNPLSTLSGWSALEBDLYLKIEIFGKNS 530
QY 452 CAIAR--LIGTYCRQVVEFRVYESIAPAB-----AEDVDTPPRK 491
DB 531 CLIANRLSLGLTCTCEIVANNVYNNNAAMAKRLLINKSISGFAENBDYMEQDMARTHI 590
QY 492 KKRKR-----LW--AAHCKRIQLKQDSSNHYNTYOPCPHROPDSSCPCVIAQNF 543
DB 591 YRRGRNRRLKTYWKSAGH-PTVRKRTDGGKCYTQYSFC-AQQMGCGDCCPAUKGTCC 648
QY 544 EKPCCSSRCNRPFGCFC-KAQCNTKQPCVLAVBEQDPLCLTGAADHDS----- 596
DB 649 EKYCGSSCKNKKFGRCAHSQCKSRQCPCPAASRECDPDVCRNC-----NVSAGDSL 703
QY 597 -----KNVSCNCSIQSGSKHLLAPSDVAGMGIFIKDPVQKNEFISEYCGEILISOD 649
DB 704 GRLPARGDYOQGNMKLLKQOQRIILGRSDVAGMGAFIKPNVKNNDYLGEYIGELISHK 763
QY 650 EADRRGKYDYKMGSPFLNANDPVVDATRGKNIKIRFANHSVNPNCYAKVMVNGDHRIG 709
DB 764 EADKKGKITYDRANSEFLPDNDQVYLDAYRKGDKIKFANHSNPNCYAKVMVAGDHRVG 823
QY 710 IFAKRAIQTGEELFPDYRYSQADA 733
DB 824 IYAKHEILASEELFYDYRIGPDQA 847

RESULT 8
US-11-330-403-9237
; Sequence 9237, Application US/11330403
; Publication No. US20060159563A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53629)B
; CURRENT APPLICATION NUMBER: US/11/330,403
; NUMBER OF SEQ ID NOS: 2006-01-12
; NUMBER OF SEQ ID NOS: 19250
; SEQ ID NO 9237
; LENGTH: 895
; TYPE: PRF
; ORGANISM: Oryza sativa (indica cultivar-group)
US-11-330-403-9237

Query Match
Best Local Similarity 19.0%; Score 769; DB 7; Length 895;
Matches 220; Conservative 94; Mismatches 244; Indels 248; Gaps 26;

QY 95 VIPLKTLNVAASVPLTMSWSPLOON-FWVEDETVLHNIPLYMGDEVLDQDGTFTBELLINX 153
DB 144 VVQLAKLFPVNDKIPYTTWIFLDKNQRMADQVLCRRRIY-----Y 184
QY 154 DKGTHGDECGFINDERIFVELYNALGOYNDDDDDGDDPPERERE-----EKQ---KOLE 204
DB 185 DPTVNEALIC-----SESDDVBEPEBEKRVFTGEBDQLWKATQ 224
QY 205 DHRDKESRPRKFPSSDKIFEALISWFPDGTAELEKRYELTEQQLPGALPPECPTNI 264
DB 225 DHGLSR-----VANVLCQYD-ATPSIEERSEVLFKRY---EKQSSSYET 268
QY 265 DGPNAKVQBEQSLHSFHTLFCRCFCFYKDCFLH-----PFAATPNTYKRNKTETALDNK 318
DB 269 DFOQLFKTMDVALDSFNLFCRCCLVDFCDLHGCSQNLVFPCEKQYSGHGLD---NKR 325
QY 319 PCGPOCY-----OHLBKAKEFAAATLARIKTPP-----KRPGRRR 355
DB 326 PCGQRYRARRREVYQDTCNDRNACTTYNTDSRSSSLKVSATILSE----- 371
QY 356 GRLPNNSRR-----STPTINLVLESKOTDS----- 381
DB 372 ---SBDSDNRDNDKISIVETSRSKITNSRYADKSVTPPPGDASSETENVSDDMPLRLTG 428

```





QY 644 EIIISODEADRGKVDKWC-----SPLFNLNDPVYDATKGRKIRPANSVNPNCYAKVM 700  
 Db 117 EIVISWEARRRAQAET-OCLDQAYITINMADESIDATKGLARFINHSCQPCETKRM 175  
 QY 701 MVNGDHRIIGIFAKRAIQTGEELFFDYRY 728  
 Db 176 NVLGEVRVGI FAKONI PFTELSYDYNF 203

RESULT 12  
 US-10-953-349-35694  
 ; Sequence 35694, Application US/10953349  
 ; Publication No. US20060107345A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALEXANDROV, Nickolai et al.  
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
 ; FILE REFERENCE: 2750-1579PUS2  
 ; CURRENT APPLICATION NUMBER: US/10/953,349  
 ; CURRENT FILING DATE: 2004-09-30  
 ; NUMBER OF SEQ ID NOS: 40252  
 ; SOFTWARE: Patent in version 3.3  
 ; SEQ ID NO 35694  
 ; LENGTH: 551  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays subsp. mays  
 US-10-953-349-35694

Query Match 7.3%; Score 297; DB 6; Length 551;  
 Best Local Similarity 33.7%; Pred. No. 5e-11;  
 Matches 70; Conservative 28; Mismatches 84; Indels 26; Gaps 6;

QY 536 CVIAQNFCE-----KFCQCSSECONRFPGRCKCAQCNCKQPC-----YLAVERCDPD 583  
 Db 45 CVEAPSYIHIDSNASVYRHKSGQTEBDIAVCEQYDILBPSPCGDRCLNLTNTECTPG 104  
 QY 584 LCLTGAADHMSKNVSCNCSIQGSKGHLILASPDVANGIFIKDPVQKNEPISYKCG 643  
 Db 105 YC-RCG-----VYCNRQRFQKQYARTRLVRTGSGWGLVADENITMAQOFVLEYCG 154  
 QY 644 EIIISODEADRGKVDKWC-----SPLFNLNDPVYDATKGRKIRPANSVNPNCYAKVM 700  
 Db 155 EIVISWEARRRAQAET-OCLDQAYITINMADESIDATKGLARFINHSCQPCETKRM 213  
 QY 701 MVNGDHRIIGIFAKRAIQTGEELFFDYRY 728  
 Db 214 NVLGEVRVGI FAKONI PFTELSYDYNF 241

RESULT 13  
 US-11-330-403-11703  
 ; Sequence 11703, Application US/11330403  
 ; Publication No. US20060159563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53629)B  
 ; CURRENT APPLICATION NUMBER: US/11/330,403  
 ; CURRENT FILING DATE: 2006-01-12  
 ; NUMBER OF SEQ ID NOS: 19250  
 ; SEQ ID NO 11703  
 ; LENGTH: 425  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 US-11-330-403-11703

Query Match 7.2%; Score 291; DB 7; Length 425;  
 Best Local Similarity 28.4%; Pred. No. 8.6e-11;  
 Matches 80; Conservative 33; Mismatches 95; Indels 74; Gaps 10;

Db 150 EYVFTLAKIDEVFCRLPVVYVNEDEPKIDSTIRDTFAAVCKEPPHFFIRRVYLIRK---- 205  
 QY 516 HVMYQPDHRRQPDSSCPVIAQNFCEKQCQSSSECONRFPGRCKCAQCNCKQPCYL 575  
 Db 206 -----KRPSRAEACCTNCSADSTCKDC-----EGRGLYMSC-----SKNCHC-- 244  
 QY 576 AVRECDPDLCTGGAADHMSKNVSCNCSIQGSKGHLILASPDVANGIFIKDPVQKN 635  
 Db 245 -----SDMCTNKPFRKOKKI KAVTKRC-----GMGALISLEPLKRG 280  
 QY 636 EPISEYCEIISODEAD-----RGKVYDK--YMCSPFLNNDPVYDATKGRKIRF 686  
 Db 281 DFIIEYGEVINDATCEQRLMDMKRG--DKNPFYMC-----EISKDFITIDATFKGNTSRF 333  
 QY 687 ANHSVNPNCYAKVMVNGDHRIGIFAKRAIQTGEELFFDYRY 728  
 Db 334 LNHSCDPNCKLEKQVDETRVGVASHSIQVGEHLTYDYRF 375

RESULT 14  
 US-11-330-403-9402  
 ; Sequence 9402, Application US/11330403  
 ; Publication No. US20060159563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53629)B  
 ; CURRENT APPLICATION NUMBER: US/11/330,403  
 ; CURRENT FILING DATE: 2006-01-12  
 ; NUMBER OF SEQ ID NOS: 19250  
 ; SEQ ID NO 9402  
 ; LENGTH: 369  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 US-11-330-403-9402

Query Match 7.1%; Score 287.5; DB 7; Length 369;  
 Best Local Similarity 22.4%; Pred. No. 1.2e-10;  
 Matches 108; Conservative 61; Mismatches 146; Indels 167; Gaps 20;

QY 287 RCFKYDCLFHPHATPNTYRKNTETALDNKP-----CGPOCYOLBGAKEFAALTAER 342  
 Db 15 RRCRL-----RQRRRRRGKQNPVPATARPSPRWADGAAPASRALAR 59  
 QY 343 I-----KTPPKRFG--GRGRGLPNNSSRSTPIITVLESKDTDSDEAGETGGENND 394  
 Db 60 ARAPLSPRAPRAAGAKSRPPRPSSTVAT-----APAHPSGRPFE 103  
 QY 395 KEEBEKDETSSSEANSRCQTPIMKKNIBPENVENWGSABSMFVLIGTYDNFCAL 454  
 Db 104 EYKEMKAKKALGVPAKCELP-----LTGT----- 131  
 QY 455 ARLIGTKCQVYEFVRESSIIAPAPADVDTPPKKKRKLRLMAHCRKIQLKDGSS 514  
 Db 132 KPAVBCRLCSKTIH-----PGEET-----KCSVSRGCEL 160  
 QY 515 NHVINYQPDHROPDCSSCPVIAQNFCEKQCQSSSECONRFPGRCKCAQCNCKQPCY 574  
 Db 161 FHL-----TCAADPTNFIAR-----SPKCPQHEVFCRLPVVYVNDPFXID 201  
 QY 575 LAVRE-----CDP-----DLCLTGAADHMSKNVSCN--CSIQGSK--KHLL 615  
 Db 202 STIRDTFAVCKEPPHFFIRRSILKGLYMSC-----SKNCHCSDMCTNKPFRKOKKIK 254  
 QY 616 LAPSDVANGIFIKDPVQKNEPISYCEIISODEAD-----RGKVYDK--YMCSPFL 666  
 Db 255 AVKTRCGMGWALISLEPLEKGDPIIEYGEVINDATCEQRLMDMKRG--DKNPFYMC-- 308  
 QY 667 FNLNDPVYDATKGRKIRPANSVNPNCYAKVMVNGDHRIGIFAKRAIQTGEELFFDY 726  
 Db 309 -EISKDFITIDATFKGNTSRFINHSCDPNCKLEKQVDETRVGVASHSIQVGEHLTYDY 367  
 QY 727 RY 728

Db 368 RW 369

RESULT 15  
US-11-330-403-2871  
; Sequence 2871, Application US/11330403  
; Publication No. US20060159563A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53629)B  
; CURRENT APPLICATION NUMBER: US/11/330,403  
; CURRENT FILING DATE: 2006-01-12  
; NUMBER OF SEQ ID NOS: 19250  
; SEQ ID NO 2871  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: ASPERGILLUS NIDULANS FGSC A4  
US-11-330-403-2871

Query Match 7.1%; Score 286.5; DB 7; Length 980;  
Best Local Similarity 23.4%; Pred. No. 4.5e-10;  
Matches 106; Conservative 67; Mismatches 157; Indels 123; Gaps 19;

QY 325 YOHLEGAKEPAA-ALTAERIK-----TPPKRPGRRGRGLPNNSSRPSTPTI----- 370  
DB 4 HDNADRGSEFVADAVTANKLEQSENNTDAPILNGG--AAKKPDSKAASPEPLIKDERAS 61  
QY 371 -NVLBSKDTSDREAGTGTGGENDKE--EBEKDETSSEANSRCQPIKMKNTIEPP 427  
DB 62 STFMKSRSSTPSSRTPLKKEHSDIEDIOBKRGDASGTEKVGGIS--VKMEPG-QPP 118  
QY 428 ENVEWSEGAEMFRLIGTYDNFCATRLIGTKTCQVYEFVYESSIIAPAPAE-DVD 486  
DB 119 K-----LAK-----SSQKVPRPPQLFLD 138  
QY 487 TPKRKKRRLMAAHCRIQLKKDSSNHVYVQPCDHPRPCDSCPCVI----- 538  
DB 139 LPDSTBEAQKTFEVIETQYANKVGYTEHA--MECD-----CAEHWLVVVLAPSPSF 190  
QY 539 ---AQN---FCEKFCQCSSECONRPFPGCRCAQCNTKQPCYLAVRECDPDLCTCGAAD 592  
DB 191 RVPSQNPASSTTRACGEDSDCINRATKIECMGDCG-----CGPD----- 229  
QY 593 HMDSNVSCKNCSTIORGSKHLLAPSDVAGWGIPTKDVOXNBFISEYCEIISQDEAD 652  
DB 230 -----CONORFQRRRYANAVIKTEKKGIGLAEEDLRPHQTFIYVGEVINEGPFH 281  
QY 653 RRGKVVYD---KYMCSFLNLNDFVVDATRKGNKIRPANHSVNPNCYAKVMVNGDRI 708  
DB 282 RRMQYDAEGIKHF--YFMSLSKSEFVDATKGNLGRFCNHSNPNCTVVDKMWVGEKLRM 339  
QY 709 GIPAKRAIQTGHELEFPDY---RYSQADALKTYG 738  
DB 340 GIPAEKHIQAGEKELVFNYVDRYGADPQPCYCG 372

Search completed: August 11, 2006, 22:50:45  
Job time : 28.2318 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2006, 11:14:57 ; Search time 12863 Seconds

(without alignments)  
11140.960 Million cell updates/sec

Title: US-10-773-302-1\_COPY\_90\_2330

Perfect score: 2241

Sequence: 1 atggcgccagctgggaagaa.....gagaatggaatcccttga 2241

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_scs:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vl:\*  
11: gb\_ov:\*  
12: gb\_hcg:\*  
13: gb\_in:\*  
14: gb\_om:\*  
15: gb\_da:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2241	100.0	2511	2	CQ714720 Sequence
2	2241	100.0	2600	2	A58329 Sequence 1
3	2241	100.0	2600	2	AR473734 Sequence
4	2241	100.0	2600	2	AX333675 Sequence
5	2241	100.0	2600	5	HSU61145 Human enhan
6	2236.2	99.8	2558	2	BD186247 STX16 act
7	2236.2	99.8	2576	2	AX821950 Sequence
8	2236.2	99.8	2576	5	HS2ESTENH
9	2229	99.5	2512	2	DD208658 Expressio
10	2216	98.9	2619	5	BC010858 Homo sapi
11	2204.2	98.4	2619	5	AB179385 Macaca fa
12	1993.8	89.0	2476	2	AR178682 Sequence
13	1993.8	89.0	2476	2	AX747657 Sequence
14	1993.8	89.0	2476	5	AK092676 Homo sapi
15	1973.8	88.1	2600	6	AK220174 Mus muscu
16	1973.8	88.1	2653	6	MMU52951 Mus muscu
17	1969	87.9	2574	6	BC003772 Mus muscu
18	1969	87.9	2595	6	BC016391 Mus muscu

19	1883.2	84.0	2617	6	BC079538 Mus muscu
20	1881.4	84.0	2546	5	AB168941 Macaca fa
21	1847.4	82.4	2665	5	HSU52965 Human putat
22	1700	75.9	1960	2	BD186308 STX16 act
23	1530.6	68.3	2658	11	BC084193 Xenopus t
24	1527.4	68.2	2631	11	CR855647 Xenopus t
25	1521	67.9	2772	11	AF351126 Xenopus t
26	1506.6	67.2	2636	11	BC097526 Xenopus t
27	1201	53.6	3641	5	AB208895 Homo sapi
28	1090.8	48.7	3394	11	AB195561 Homo sapi
29	859	38.3	859	5	AY519465 Homo sapi
30	840.4	37.5	2244	8	BT009782 Homo sapi
31	840.4	37.5	2244	8	AY889044 Synthetic
32	840.4	37.5	2594	2	CG215926 Sequence
33	840.4	37.5	2594	2	AX821949 Sequence
34	840.4	37.5	2594	5	HSU50315 Human enhan
35	840.4	37.5	4606	5	AB002386 Homo sapi
36	840.4	37.5	4640	5	BC015882 Homo sapi
37	828.8	37.0	2253	6	AF483490 Mus muscu
38	828.8	37.0	2253	6	AF483491 Mus muscu
39	828.8	37.0	3970	6	MMU60453 Mus muscu
40	828.8	37.0	4185	6	AB004817 Mus muscu
41	827.2	36.9	4179	6	BC007135 Mus muscu
42	818.6	36.5	4147	6	AK129004 Mus muscu
43	814.6	36.3	4639	5	HSN804470 Homo sapi
44	738	32.9	1889	11	BC095094 Dario rex
45	710.6	31.7	2749	5	AB004818 Homo sapi

## ALIGNMENTS

RESULT 1	CQ714720	2511 bp	DNA	linear	PAT 03-FEB-2004
LOCUS	Sequence 654 from Patent WO02068579.				
DEFINITION	CQ714720				
ACCESSION	CQ714720				
VERSION	CQ714720.1 GI:42275577				
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;				
	Hominidae; Homo.				
REFERENCES					
AUTHORS	Venter, C.J., Adams, M.C., Li, P.W. and Myers, B.W.				
TITLE	Kits, such as nucleic acid arrays, comprising a majority of				
	humanexons or transcripts, for detecting expression and other uses				
	thereof				
JOURNAL	Patent: WO 02068579-A 654 06-SEP-2002;				
	PE Corporation (NY) (US)				
FEATURES					
source	1..2511				
	Location/Qualifiers				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
Query Match	100.0%; Score 2241; DB 2; Length 2511;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 2241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 ATGGCCAGACTGGAGAGAAATCTGAGAGGACCAAGTTTGGCGGAGCGTCTAA	60			
DB	10 ATGGCCAGACTGGAGAGAAATCTGAGAGGACCAAGTTTGGCGGAGCGTCTAA	69			
QY	61 TCAGAGTCATGCGCATGAGACAGCTCAAGAGTTCAGACGAGCTGAAGTAAGAGT	120			
DB	70 TCAGAGTCATGCGCATGAGACAGCTCAAGAGTTCAGACGAGCTGAAGTAAGAGT	129			
QY	121 ATGTTTACCTTCAATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGATG	180			
DB	130 ATGTTTACCTTCAATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGATG	189			

[illegible]

QY	1261	AAGCAAAATATTTGAACCTCCGAGAAAGTGGAGTGGATGGTGAAGACCTCAATGTTT	1320
Db	1270	AAGCAAAATATTTGAACCTCCGAGAAAGTGGAGTGGATGGTGAAGACCTCAATGTTT	1329
QY	1321	AGAGTCTCATTTGGACATTACTATGACAAATTTCTGTGCCATTTGCTAGGTTAATTGGAC	1380
Db	1330	AGAGTCTCATTTGGACATTACTATGACAAATTTCTGTGCCATTTGCTAGGTTAATTGGAC	1389
QY	1381	AAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAAGAAATCTAGCATCATAGCTCCAGCT	1440
Db	1390	AAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAAGAAATCTAGCATCATAGCTCCAGCT	1449
QY	1441	CCCGCTGAGGATGTGATATCTCTCCCAAGGAAAAAGAGAGAAACACCGGTGTGGGCT	1500
Db	1450	CCCGCTGAGGATGTGATATCTCTCCCAAGGAAAAAGAGAGAAACACCGGTGTGGGCT	1509
QY	1501	GCACACTGCAGAAAAGATPACAGCTGAAAAAGACCGGCTCTCTTAAACAATGTTTACATAT	1560
Db	1510	GCACACTGCAGAAAAGATPACAGCTGAAAAAGACCGGCTCTCTTAAACAATGTTTACATAT	1569
QY	1561	CAACCCGTGTATTCATCCACGGCAGCCCTTGTGACAGTTCGGCCCTTGTGTATGACAA	1620
Db	1570	CAACCCGTGTATTCATCCACGGCAGCCCTTGTGACAGTTCGGCCCTTGTGTATGACAA	1629
QY	1621	AAATTTTGTGAAAAAGTTTGTCAATGTATGTTACAGAGTGCAMAAACCGCTTTCGGGATGC	1680
Db	1630	AAATTTTGTGAAAAAGTTTGTCAATGTATGTTACAGAGTGCAMAAACCGCTTTCGGGATGC	1689
QY	1681	CGCTGCAGAGCACATGTGCACACCAAGCAGTGCCTGTCTACTGTGGCTGTCCAGAGTGT	1740
Db	1690	CGCTGCAGAGCACATGTGCACACCAAGCAGTGCCTGTCTACTGTGGCTGTCCAGAGTGT	1749
QY	1741	GACCTTACACCTGTCTTACTTGTGTGAGACCGCTGCACATTTGGGACATTAATAAATGTGTCC	1800
Db	1750	GACCTTACACCTGTCTTACTTGTGTGAGACCGCTGCACATTTGGGACATTAATAAATGTGTCC	1809
QY	1801	TGCAGAAAGACTGCAGTATTCAGCGGGGCTCCAAAAAGCATGTATTTGCTGGACCATCTGAC	1860
Db	1810	TGCAGAAAGACTGCAGTATTCAGCGGGGCTCCAAAAAGCATGTATTTGCTGGACCATCTGAC	1869
QY	1861	GTGGCAGGCTGTGGGGGATTTTTTATCAAAGATCTGTGCAGAAAAATGAATTCATCTCAGAA	1920
Db	1870	GTGGCAGGCTGTGGGGGATTTTTTATCAAAGATCTGTGTGAGAAAAATGAATTCATCTCAGAA	1929
QY	1921	TACTGTGGAAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGGGAAAGTATGATPAA	1980
Db	1930	TACTGTGGAAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGGGAAAGTATGATPAA	1989
QY	1981	TACATGTGCAGCTTTCGTGTCAACTTGAACAATGATTTTGTGTGATGCACCCGCAAG	2040
Db	1990	TACATGTGCAGCTTTCGTGTCAACTTGAACAATGATTTTGTGTGATGCACCCGCAAG	2049
QY	2041	GCTTAAACAAATTCGTTTTCGAAATCAATTCGTGTAAATCAAACTGTCTATGCAAAAGTTATG	2100
Db	2050	GCTTAAACAAATTCGTTTTCGAAATCAATTCGTGTAAATCAAACTGTCTATGCAAAAGTTATG	2109
QY	2101	ATGTGTTAACGGTGTACACAGATAGGATTTTTTGTCCAAAGAGAGCATCCAGACTGGGAA	2160
Db	2110	ATGTGTTAACGGTGTACACAGATAGGATTTTTTGTCCAAAGAGAGCATCCAGACTGGGAA	2169
QY	2161	GAGCTGTTTTTGTATTAACAATACAGCCACGAGCTGATGCCCTGAAGTATGTGTGGCATGAA	2220
Db	2170	GAGCTGTTTTTGTATTAACAATACAGCCACGAGCTGATGCCCTGAAGTATGTGTGGCATGAA	2229
QY	2221	AGAGAAATGAAAAATCCCTTGA 2241	
Db	2230	AGAGAAATGAAAAATCCCTTGA 2250	

RESULT 2			
A58329			
LOCUS	2600 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO9635784.		PAT 05-MAR-1998

ACCESSION A58329 GI:3713991  
VERSION A58329.1  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE 1  
AUTHORS Jernwein,T. and Laible,G.  
TITLE CHROMATIN-REGULATOR GENES  
JOURNAL Patent: WO 9635784-A 1 14-NOV-1996;  
BOEHRINGER INGENIEUR INT (DE)  
COMMENT Other publication DE 19516776 961114.  
FEATURES  
source  
1..2600  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/cell\_type="B-CELL"  
1..89  
90..2330  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAA03482.1"  
/db\_xref="GI:3713992"  
/translation="MGQTGKSEKGPVCKRKYKSEYHRLQLKFRPADRYKSPSS  
NQKILERTILNQEWKQRIQPVHILTVSSLSKGRVCSVSDLDPTQVPLKTLN  
AVASVPIVMSPLQONFVDEDETVLHNPYMGDEVLDQDGFIELIKNYDGKXHG  
RECGFINDEIPELVNALQYNDDDDDGDDPEREKKDLBHDKESRPERK  
PEDKIFEAISMPKGTABELKEKYKELTEQOLPGALPEPCTPNIIDPNMAKSVREO  
SLASFHTLFCRCPCFYDCETLHPHATPNTYKNTETLIDNKPCGPOCOHLGAKER  
AAALYARIKTPKPRKPRGRRLRNNSRSTPTIINLESQDTSDREAGETGSEN  
NDKEBEKDETSSESEANSRCQPIKMPNIEPPEVMSGAESMRLALGYTDN  
FCALRIIGTKTCROVETFRVRESIIAPAPAEVDVTPRKKRKHRLMAVLCRIOL  
KDGSSNHYNTQPCDHPROPDSCPCVIAQNPCEKFCQSSSECONFPGRCAQC  
NTKOCPCYLAVERCDLCTGGAADHMSKNVSKNCSIQSGSKHLLIABSDVAG  
GIFIDPQKNEFISEYGERIISODEADPRGVYDKMCSPLPNNNDPVDAITKYN  
KIRPANHVNPCYAKVMVNGDHRIGIPAKAIDTGEELFPDYRISQADALKYVIGIE  
REMEIP"  
2331..2600  
ORIGIN  
3'UTR  
Query Match 100.0%; Score 2241; DB 2; Length 2600;  
Beet Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGCGAGACTGGGAAAGAAATCTGAGAGGAGACAGTTGTTGGCGAAAGCGTGTAA 60  
DB 90 ATGGGCGAGACTGGGAAAGAAATCTGAGAGGAGACAGTTGTTGGCGAAAGCGTGTAA 149  
QY 61 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAGACGACTGATGAAGTAAAGCT 120  
DB 150 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAGACGACTGATGAAGTAAAGCT 209  
QY 121 ATGTTTAACTTCAATCTGAGAAATTTTGGAAAGAAAGGAAATCTTAAACCAAGATGG 180  
DB 210 ATGTTTAACTTCAATCTGAGAAATTTTGGAAAGAAAGGAAATCTTAAACCAAGATGG 269  
QY 181 AAACGCGAAGATCAGCGTGTGACATCTGACTTGTGAGGCTCATTTGGCGGAGACT 240  
DB 270 AAACGCGAAGATCAGCGTGTGACATCTGACTTGTGAGGCTCATTTGGCGGAGACT 329  
QY 241 AGGAGTGTTCGGTGCACAGTACATTTTCCAAACAAGTCAATCCATTAAAGACT 300  
DB 330 AGGAGTGTTCGGTGCACAGTACATTTTCCAAACAAGTCAATCCATTAAAGACT 389  
QY 301 CTGAATGCACTTCTTCACTACCAATATATTTCTTGTCTCCCTTACAGAGAAATTTT 360  
DB 390 CTGAATGCACTTCTTCACTACCAATATATTTCTTGTCTCCCTTACAGAGAAATTTT 449  
QY 361 ATGTGGAAGATGAAGTCTTTTACATTAATCTCTTATATGAGAGATGAAGTTTAACT 420

DB 450 ATGTGGAAGATGAAGTCTTTTACATTAATCTCTTATATGAGAGATGAAGTTTAACT 509  
QY 421 CAGAGTGTATCTTCAATGAAGAACTAATPAAAAATTTATGAGGAAAGTACACGGGAT 480  
DB 510 CAGAGTGTATCTTCAATGAAGAACTAATPAAAAATTTATGAGGAAAGTACACGGGAT 569  
QY 481 AGAGATGTGGTTTATTAATGATGAAGAAATTTTGTGAGTTGATGATGATGATGATGAT 540  
DB 570 AGAGATGTGGTTTATTAATGATGAAGAAATTTTGTGAGTTGATGATGATGATGATGATGAT 629  
QY 541 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 630 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 601 AAAGATCTGAGAGATCAACGAGATGATTAAGAAACCGCCCACTTCGAAATTTCTCTCT 660  
DB 690 AAAGATCTGAGAGATCAACGAGATGATTAAGAAACCGCCCACTTCGAAATTTCTCTCT 749  
QY 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGCGACAGAGAAAGACTA 720  
DB 750 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGCGACAGAGAAAGACTA 809  
QY 721 AAGAAAAATATTAAGAACTCAACGAGAGAGCTCCAGGCGACTTCTCTGATATGT 780  
DB 810 AAGAAAAATATTAAGAACTCAACGAGAGAGCTCCAGGCGACTTCTCTGATATGT 869  
QY 781 ACCCCCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
DB 870 ACCCCCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929  
QY 841 TTTTATGAGCTTTTCTGATGAGGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 900  
DB 930 TTTTATGAGCTTTTCTGATGAGGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 989  
QY 901 GCAACCACTAATTAAGGAGAGAAACAGAAACAGCTTGAACAACCACTTGT 960  
DB 990 GCAACCACTAATTAAGGAGAGAAACAGAAACAGCTTGAACAACCACTTGT 1049  
QY 961 GCAACCACTAATTAAGGAGAGAAACAGAAACAGCTTGAACAACCACTTGT 1020  
DB 1050 GCAACCACTAATTAAGGAGAGAAACAGAAACAGCTTGAACAACCACTTGT 1109  
QY 1021 GAGCGGATTAAGAACCCCAACCAAACTGTCAGAGGCGCAGAGAGAGAGCGCTTCCCAT 1080  
DB 1110 GAGCGGATTAAGAACCCCAACCAAACTGTCAGAGGCGCAGAGAGAGAGCGCTTCCCAT 1169  
QY 1081 AACAGTACAGGCGCCAGACCCCACTTAATGTGCTGAATCAAGATACAGACGT 1140  
DB 1170 AACAGTACAGGCGCCAGACCCCACTTAATGTGCTGAATCAAGATACAGACGT 1229  
QY 1141 GATAGGAG 1200  
DB 1230 GATAGGAG 1289  
QY 1201 AAAGATGAACCTGAGAGCTCTGAGAGCAATCTCGGTGCAAAACCAATTAAGATG 1260  
DB 1290 AAAGATGAACCTGAGAGCTCTGAGAGCAATCTCGGTGCAAAACCAATTAAGATG 1349  
QY 1261 AAGCAATTAATGAACCTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1350 AAGCAATTAATGAACCTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409  
QY 1321 AGAGTCTCATTTGAGACTTACTATGACAAATTTCTGTGCAATTTCTAGTTAATGGAGCC 1380  
DB 1410 AGAGTCTCATTTGAGACTTACTATGACAAATTTCTGTGCAATTTCTAGTTAATGGAGCC 1469  
QY 1381 AAAACATGTAGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1470 AAAACATGTAGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529  
QY 1441 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
DB 1530 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589

Qy	1501	GCACACTGCAGAAAGATTCACACTGAAAAAGGACGGCTCTCTTAACCATATGTTTAAACAATAT	1560
Db	1590	GCACACTGCAGAAAGATTCACACTGAAAAAGGACGGCTCTCTTAACCATATGTTTAAACAATAT	1649
Qy	1561	CAACCCCTGTGATTCATCCACGGCAGCCTTGTCAGACTTCGTCGCCCTGTGTGATAGCAAA	1620
Db	1550	CAACCCCTGTGATTCATCCACGGCAGCCTTGTCAGACTTCGTCGCCCTGTGTGATAGCAAA	1709
Qy	1621	AATTTTGTGAAAAAGTTTGTGCATGTATGTTTCAGAGTGTCAAAAACGGCTTTCGGGATGC	1680
Db	1710	AATTTTGTGAAAAAGTTTGTGCATGTATGTTTCAGAGTGTCAAAAACGGCTTTCGGGATGC	1769
Qy	1681	CGCTGCAGAAAGCAGGTGCAACCAAGCAAGGAGTGCCTGCTCACTGGCTGTCCGAGAGTGT	1740
Db	1770	CGCTGCAGAAAGCAGGTGCAACCAAGCAAGGAGTGCCTGCTCACTGGCTGTCCGAGAGTGT	1829
Qy	1741	GACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATTGGGACAGTAAAAATGTGTCC	1800
Db	1830	GACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATTGGGACAGTAAAAATGTGTCC	1889
Qy	1801	TGCAGAAACTGCAAGTATTCAGCGGGGGCTTCCAAAAAGCATCTATTTGCTGTGGCACTTGAC	1860
Db	1890	TGCAGAAACTGCAAGTATTCAGCGGGGGCTTCCAAAAAGCATCTATTTGCTGTGGCACTTGAC	1949
Qy	1861	GTGGCAGGCTGGGGGATTTTTCATCAAGATCTGTGSCAGAAAAATGAAATTCATCTCAGAA	1920
Db	1950	GTGGCAGGCTGGGGGATTTTTCATCAAGATCTGTGSCAGAAAAATGAAATTCATCTCAGAA	2009
Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAACTGACAGAAAGGAAAAAGTATGATTA	1980
Db	2010	TACTGTGAGAGATTAATTTCTCAAGATGAACTGACAGAAAGGAAAAAGTATGATTA	2069
Qy	1991	TACATGTGCAGCTTTCTGTGTCAACTTGAAACATGATTTTGTGTGAGATGCAACCCGCAAG	2040
Db	2070	TACATGTGCAGCTTTCTGTGTCAACTTGAAACATGATTTTGTGTGAGATGCAACCCGCAAG	2129
Qy	2041	GGTAAACAATAATCGTTTTTGCAAAATCAATTCGGTAAATCCAACTGCTATGCAAAAGTTATG	2100
Db	2130	GGTAAACAATAATCGTTTTTGCAAAATCAATTCGGTAAATCCAACTGCTATGCAAAAGTTATG	2189
Qy	2101	ATGTGTAAACGGTGATCACAGATAGGTAATTTTGTGCCAAGAGAGCCATTCAGACTGGCGAA	2160
Db	2190	ATGTGTAAACGGTGATCACAGATAGGTAATTTTGTGCCAAGAGAGCCATTCAGACTGGCGAA	2249
Qy	2161	GAGCTGTTTTTTGATTAACAGATPACAGCCAGGCTGATGCCCTGGAATATGTGGGATTCGAA	2220
Db	2250	GAGCTGTTTTTTGATTAACAGATPACAGCCAGGCTGATGCCCTGGAATATGTGGGATTCGAA	2309
Qy	2221	AGAGAAATGAGAAATCCCTTGA 2241	
Db	2310	AGAGAAATGAGAAATCCCTTGA 2310	
RESULT 3			
LOCUS	AR473734	2600 bp	DNA
DEFINITION	Sequence 1 from patent US 668583.		linear
LOCUS	AR473734		
VERSION	AR473734.1	GI:42712196	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2600)		
AUTHORS	Jennwein, R., Labile, G., O'Carroll, D., Eisenhaber, F. and Rea, S.		
TITLE	Chromatin regulator genes		
JOURNAL	Patent: US 668583-A 1 10-FEB-2004;		
	Boehringer Ingelheim International GmbH;;		
FEATURES	DEX;	Location/Qualifiers	
SOURCE	1.. 2600		
	/organism="unknown"		

ORIGIN	/mol_type="genomic DNA"
Query Match	100.0%; Score 2241; DB 2; Length 2600;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ATGGCCGACCTGGGAAGAATCTGAGAAAGGCACAATTGTGTGGCGGAAGCGGTAAAA 60
Db	90 ATGGGCCAAGACTGGGAAGAAATCTGAGAAAGGCACAATTGTGTGGCGGAAGCGGTAAAA 149
OY	61 TCAGATGACATGCGACTGAGACAGCTCAAGAGTTTCAGACGACTGATGAAATGTAAGT 120
Db	150 TCAAGATGACATGCGACTGAGACAGCTCAAGAGTTTCAGACGACTGATGAAATGTAAGT 209
OY	121 ATGTTTAGTTCCAATCGTCAGAAAATTTTGAAAAGAACGGAATCTTTAAACAAAGATGG 180
Db	210 ATGTTTAGTTCCAATCGTCAGAAAATTTTGAAAAGAACGGAATCTTTAAACAAAGATGG 269
OY	181 AAACAGCAGAGAGATACAGCTGTGCAATCTCTGACTTCTGTAGCTCATATGGCGGGACT 240
Db	270 AAACAGCAGAGAGATACAGCTGTGCAATCTCTGACTTCTGTAGCTCATATGGCGGGACT 329
OY	241 AGGAGGTTCGGGAGCACAGTACCTTGATTTTTCACAACAAGTCACTCCATTAAAGACT 300
Db	330 AGGAGGTTCGGGAGCACAGTACCTTGATTTTTCACAACAAGTCACTCCATTAAAGACT 389
OY	301 CTGAATCAGTGTGCTTCAGTACCATAATGTAATCTTGTGATCTCCCTCAGCAGAAATTTT 360
Db	390 CTGAATCAGTGTGCTTCAGTACCATAATGTAATCTTGTGATCTCCCTCAGCAGAAATTTT 449
OY	361 ATGCTGGAAGATGAAACTGTTTTTACATPAACAATTCCTTATATGGAAGATGAAGTTTATGAT 420
Db	450 ATGCTGGAAGATGAAACTGTTTTTACATPAACAATTCCTTATATGGAAGATGAAGTTTATGAT 509
OY	421 CAGGATGTAATCTTCAATTGAAGAACTAAATPAAAAAATTATATGAGGAAAGTACACGGGAT 480
Db	510 CAGGATGTAATCTTCAATTGAAGAACTAAATPAAAAAATTATATGAGGAAAGTACACGGGAT 569
OY	481 AGAGATATGGGTTTTATTAATGATGAATTTTGTGAGTGTGTAATGCCCTTGTGCAA 540
Db	570 AGAGATATGGGTTTTATTAATGATGAATTTTGTGAGTGTGTAATGCCCTTGTGCAA 629
OY	541 TATATATGATGATGACATGATGATGATGAGACGATCTCTGAAGAAAGAGAAAGACAG 600
Db	630 TATATATGATGATGACATGATGATGATGAGACGATCTCTGAAGAAAGAGAAAGACAG 689
OY	601 AAAATATCGAGAGATCAACCGAGATGAAAGAAAGCGCCCACTCGGAAATTTCTCTCT 660
Db	690 AAAATATCGAGAGATCAACCGAGATGAAAGAAAGCGCCCACTCGGAAATTTCTCTCT 749
OY	661 GATAAAAATTTTGAAGCAATTTCTCAATGTTTTCAGATTAAGGACAGACAGAAAGAACTA 720
Db	750 GATAAAAATTTTGAAGCAATTTCTCAATGTTTTCAGATTAAGGACAGACAGAAAGAACTA 809
OY	721 AAGAAAAATATTAAGAACTCACCGGAAACAGACGCTCCAGGCGCACTTCTCTGAAATGT 780
Db	810 AAGAAAAATATTAAGAACTCACCGGAAACAGACGCTCCAGGCGCACTTCTCTGAAATGT 869
OY	781 ACCCCCAACATGATGGAACAAATGCTAAATCTGTTACAGAGAGCAAAAGCTTACATCC 840
Db	870 ACCCCCAACATGATGGAACAAATGCTAAATCTGTTACAGAGAGCAAAAGCTTACATCC 929
OY	841 TTTCATACGCTTTCTGTAGGCGAATGTTTTAAATATGATGCTCTCTCAACATCTTTTCAT 900
Db	930 TTTCATACGCTTTCTGTAGGCGAATGTTTTAAATATGATGCTCTCTCTCAACATCTTTTCAT 989
OY	901 GCACACCCCAACACTTATTAAGCGGAGAAACACAGAAACAGCTTGAACAACAACCTTGT 960
Db	990 GCACACCCCAACACTTATTAAGCGGAGAAACACAGAAACAGCTTGAACAACAACCTTGT 1049
OY	961 GGACCAAGTGTATCCAGCAATTTGAGAGGAGCAAAAGAGTTTGTGTGCTGTCAACGGCT 1020

```

Db      1050 GGACCAAGTGTATACAGACTTTTGAGGAGCAAAAGAGTTTGTCTGTCTCAACGGCT 1109
Qy      1021 GAGCGGATTAAGACCCCAACCAAAAGTCCAGAGAGCGCGCAAGAGAGAGCGCTTCCCAT 1080
Db      1110 GAGCGGATTAAGACCCCAACCAAAAGTCCAGAGAGCGCGCAAGAGAGAGCGCTTCCCAT 1169
Qy      1081 AACAGTAGAGGCGCCGACACCCCAACCAAAAGTCCAGAGAGCGCGCAAGAGAGAGCGCT 1140
Db      1170 AACAGTAGAGGCGCCGACACCCCAACCAAAAGTCCAGAGAGCGCGCAAGAGAGAGCGCT 1229
Qy      1141 GATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db      1230 GATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1289
Qy      1201 AAAAGTAGAACTTCCAGCTCTCTGAGCAAAATTTCTCGGTGCAAAACCAATTAAGATG 1260
Db      1230 AAAAGTAGAACTTCCAGCTCTCTGAGCAAAATTTCTCGGTGCAAAACCAATTAAGATG 1349
Qy      1261 AAGCAAAATATGAACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db      1350 AAGCAAAATATGAACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409
Qy      1321 AAGCTCTCATTTGAGCACTTACTATGACAAATTTCTGTGCAATTTGCTAGATTAAATGGAGC 1380
Db      1410 AAGCTCTCATTTGAGCACTTACTATGACAAATTTCTGTGCAATTTGCTAGATTAAATGGAGC 1469
Qy      1381 AAAACATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGACATCATAGCTCCAGCT 1440
Db      1470 AAAACATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGACATCATAGCTCCAGCT 1529
Qy      1441 CCGCGTAGAGAGTGTGATCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db      1530 CCGCGTAGAGAGTGTGATCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589
Qy      1501 GACACTGTGAGAGAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db      1590 GACACTGTGAGAGAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649
Qy      1561 CAACCTGTGATCATCCAGCGAGCGCTGTGACAGTCTGTGCTGTGTGTGTGTGTGTGTGTGT 1620
Db      1650 CAACCTGTGATCATCCAGCGAGCGCTGTGACAGTCTGTGCTGTGTGTGTGTGTGTGTGT 1709
Qy      1621 AATTTTGTGAAAGTTTGTGCAATGTATGAGTGTGAGTGTGCAAAACCGCTTCCGGAGTC 1680
Db      1710 AATTTTGTGAAAGTTTGTGCAATGTATGAGTGTGAGTGTGCAAAACCGCTTCCGGAGTC 1769
Qy      1681 CGCTGCAAGAGACAGTGTGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db      1770 CGCTGCAAGAGACAGTGTGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1829
Qy      1741 GACCTGTGATCTGTCTTACTTGTGTGAGCGCTGTGACATTTGAGACATTAATAATGTGTCTC 1800
Db      1830 GACCTGTGATCTGTCTTACTTGTGTGAGCGCTGTGACATTTGAGACATTAATAATGTGTCTC 1889
Qy      1801 TGCAAGAGACTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db      1890 TGCAAGAGACTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1949
Qy      1861 GTGAGAGGCTGTGAGAGATTTTATCAAAAGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db      1950 GTGAGAGGCTGTGAGAGATTTTATCAAAAGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAG 2009
Qy      1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db      2010 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2069
Qy      1981 TACATGTGAGAGCTTTCTGTGCAACTTGAACATGATTTGTGTGTGTGTGTGTGTGTGTGT 2040
Db      2070 TACATGTGAGAGCTTTCTGTGCAACTTGAACATGATTTGTGTGTGTGTGTGTGTGTGTGT 2129
Qy      2041 GGTAAACAAATTCGTTTGTGCAATCATTCGTGTAATCCAACTGTGTATGCAAAAGTTATG 2100
Db      2130 GGTAAACAAATTCGTTTGTGCAATCATTCGTGTAATCCAACTGTGTATGCAAAAGTTATG 2189

```

```

Qy      2101 ATGTTAAGGATACAGGATAGTATTTTGTCCAAAGAGCCATCCAGCTGGCGAA 2160
Db      2190 ATGTTAAGGATACAGGATAGTATTTTGTCCAAAGAGCCATCCAGCTGGCGAA 2249
Qy      2161 GAGCTGTGTTTGTATTAAGATACAGCAGGCTGATGCTGTGATATGTCCGATCGAA 2220
Db      2250 GAGCTGTGTTTGTATTAAGATACAGCAGGCTGATGCTGTGATATGTCCGATCGAA 2309
Qy      2221 AGAGAAATGAAATCCCTTGA 2241
Db      2310 AGAGAAATGAAATCCCTTGA 2330

RESULT 4
AX333675
LOCUS      AX333675
DEFINITION Sequence 4184 from Patent WO0194629.
ACCESSION AX333675
VERSION    AX333675.1 GI:18124394
KEYWORDS
SOURCE
ORGANISM   Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE
1
AUTHORS    Young, P.B., Augustus, M., Carter, K.C., Edner, R., Endress, G.,
            Horrikan, S., Soppet, D.R. and Weaver, Z.
TITLE       Cancer gene determination and therapeutic screening using signature
            gene sets
JOURNAL     Patent: WO 0194629-A 4184 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES
            source
            1..2600
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      100.0%; Score 2241; DB 2; Length 2600;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ATGGCCAGACTGGAGAGAAATCTGAGAGGAGCAAGTTGTCGCGAGAGCTGTAA 60
Db      90 ATGGCCAGACTGGAGAGAAATCTGAGAGGAGCAAGTTGTCGCGAGAGCTGTAA 149
Qy      61 TCAGAGTACATGCCAGCTGAGACAGCTCAAGAGTTCAAGACGATGAGAGTAAAGAT 120
Db      150 TCAGAGTACATGCCAGCTGAGACAGCTCAAGAGTTCAAGACGATGAGAGTAAAGAT 209
Qy      121 ATGTTAGTTCATTCGTCAAGAAATTTTGAAGAGAGAGAAATCTTAAACCAAGATG 180
Db      210 ATGTTAGTTCATTCGTCAAGAAATTTTGAAGAGAGAGAAATCTTAAACCAAGATG 269
Qy      181 AAACAGAGAGATACAGCTGTGTGACATCTGTGTGAGTCAATTTGCGCGGAGCT 240
Db      270 AAACAGAGAGATACAGCTGTGTGACATCTGTGTGAGTCAATTTGCGCGGAGCT 329
Qy      241 AGGAGAGTTTGGGTGACAGAGATTTTCCAAACAGATCATCCATTAAGACT 300
Db      330 AGGAGAGTTTGGGTGACAGAGATTTTCCAAACAGATCATCCATTAAGACT 389
Qy      301 CTGATGAGATGCTTCTGATACCAATATGATTTCTGCTCCCTACAGAGATTTT 360
Db      390 CTGATGAGATGCTTCTGATACCAATATGATTTCTGCTCCCTACAGAGATTTT 449
Qy      361 ATGTTGAGAGATGAACTGTTTACATTAACATTCCTTATATGAGAGATGAAAGTTTATG 420
Db      450 ATGTTGAGAGATGAACTGTTTACATTAACATTCCTTATATGAGAGATGAAAGTTTATG 509
Qy      421 CAGATGTACTTTCATTTGAAGAACTAATAAAATTAATGAGGAAAGTACAGCGGAT 480

```

Db ||||| 510 CAGATGGTACTTTCATGGAAGAACTAATAAAAAATATGATGGGAAAGTACAGGGGAT 569  
 Qy ||||| 481 AGAAGATGGGTTTATTAATGAATATTTTGGGATGGTGAATGCCCTGGTCAA 540  
 Db ||||| 570 AGAAGATGGGTTTATTAATGAATATTTTGGGATGGTGAATGCCCTGGTCAA 629  
 Qy ||||| 541 TATAATGATGATGACGATGATGATGAGACGATCTCTGAAGAAAGAAAGAGAG 600  
 Db ||||| 630 TATAATGATGATGACGATGATGATGAGACGATCTCTGAAGAAAGAAAGAGAG 689  
 Qy ||||| 601 AAAATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660  
 Db ||||| 690 AAAATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749  
 Qy ||||| 661 GATAAAATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 720  
 Db ||||| 750 GATAAAATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 809  
 Qy ||||| 721 AAGGAAATATTAAGAACTACCGAAGAGAGCTCCGAGGCGCACTTCTCTGATATG 780  
 Db ||||| 810 AAGGAAATATTAAGAACTACCGAAGAGAGCTCCGAGGCGCACTTCTCTGATATG 869  
 Qy ||||| 781 ACCCCCAATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840  
 Db ||||| 870 ACCCCCAATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 929  
 Qy ||||| 841 TTTCATAGCTTTTCTGATGGGATGTTTAAATATGATCTGCTTCTCATCTCTTTCAT 900  
 Db ||||| 930 TTTCATAGCTTTTCTGATGGGATGTTTAAATATGATCTGCTTCTCATCTCTTTCAT 989  
 Qy ||||| 901 GCAACACCCCAACTTATTAAGGGAAGAACAGAAAGCTCTAGACAAACCTTGT 960  
 Db ||||| 990 GCAACACCCCAACTTATTAAGGGAAGAACAGAAAGCTCTAGACAAACCTTGT 1049  
 Qy ||||| 961 GACCAAGCTTTTACAGACTTTTGGAGGAGCAAGAGTTTCTGCTCTCAACGCT 1020  
 Db ||||| 1050 GACCAAGCTTTTACAGACTTTTGGAGGAGCAAGAGTTTCTGCTCTCAACGCT 1109  
 Qy ||||| 1021 GAGGGAATTAAGACCCCAACAAAGCTCAAGAGGCGCAGAGAGAGAGCGGCTTCCAT 1080  
 Db ||||| 1110 GAGGGAATTAAGACCCCAACAAAGCTCAAGAGGCGCAGAGAGAGAGCGGCTTCCAT 1169  
 Qy ||||| 1081 AACAGTACAGGCCCCAGACCCCACTTAATGCTGGAATCAAGAGATACAGACGT 1140  
 Db ||||| 1170 AACAGTACAGGCCCCAGACCCCACTTAATGCTGGAATCAAGAGATACAGACGT 1229  
 Qy ||||| 1141 GATAGGGAAGCAGGAGCTGAAACGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1200  
 Db ||||| 1230 GATAGGGAAGCAGGAGCTGAAACGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1289  
 Qy ||||| 1201 AAAAGTAAACTTTCAGAGCTCCTCTGAAGCAAAATTCCTCGGTGTCAAAACCAATAAAGAT 1260  
 Db ||||| 1280 AAAAGTAAACTTTCAGAGCTCCTCTGAAGCAAAATTCCTCGGTGTCAAAACCAATAAAGAT 1349  
 Qy ||||| 1261 AAGCCAAATATTTGAACCTCTGAGATGATGAGTGAATGCTGAAAGCCCTCAATGTTT 1320  
 Db ||||| 1350 AAGCCAAATATTTGAACCTCTGAGATGATGAGTGAATGCTGAAAGCCCTCAATGTTT 1409  
 Qy ||||| 1321 AGAGTCTCATTTGGCATTTACTATGACAAATTTCTGTGCCATTTGCTAGTTTATTTGGACC 1380  
 Db ||||| 1410 AGAGTCTCATTTGGCATTTACTATGACAAATTTCTGTGCCATTTGCTAGTTTATTTGGACC 1469  
 Qy ||||| 1381 AAAACATGATGACAGGTGATGATGATTAAGATCAAAAGATCTAGCATATAGCTCAGCT 1440  
 Db ||||| 1470 AAAACATGATGACAGGTGATGATGATTAAGATCAAAAGATCTAGCATATAGCTCAGCT 1529  
 Qy ||||| 1441 CCGGCTGAGATGATGATCTCTCTCAAGAAAGAAAGAGAGAACACCGGTTTGGGCT 1500  
 Db ||||| 1530 CCGGCTGAGATGATGATCTCTCTCAAGAAAGAAAGAGAGAACACCGGTTTGGGCT 1589  
 Qy ||||| 1501 GCACATGCAAGAAATACAGCTGAAAGAGAGCGGCTCCTTAACCATGTTTACACATAT 1560

Db ||||| 1590 GCACATGCAAGAAATACAGCTGAAAGAGAGCGGCTCCTTAACCATGTTTACACTAT 1649  
 Qy ||||| 1561 CAACCCCTGATCATCCACGCGAGCCTTGTGACAGATTCTGCTTGTGATATGACAA 1620  
 Db ||||| 1650 CAACCCCTGATCATCCACGCGAGCCTTGTGACAGATTCTGCTTGTGATATGACAA 1709  
 Qy ||||| 1621 AATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAACCGCTTTCCGGATGC 1680  
 Db ||||| 1710 AATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAACCGCTTTCCGGATGC 1769  
 Qy ||||| 1681 CGCTGCAAGACAGTGCACACCAAGCAGTCCCGGTCTACTGCTGCTCCGAGATGT 1740  
 Db ||||| 1770 CGCTGCAAGACAGTGCACACCAAGCAGTCCCGGTCTACTGCTGCTCCGAGATGT 1829  
 Qy ||||| 1741 GACCTGACCTCTGCTTACTTGTGAGCCGCTGACCATTTGGGACATTAATAATGTGTCC 1800  
 Db ||||| 1830 GACCTGACCTCTGCTTACTTGTGAGCCGCTGACCATTTGGGACATTAATAATGTGTCC 1889  
 Qy ||||| 1801 TGCAAGACCTGAGATTTACGCGGGGCTCCAAAAGCATTTATGCTGGCACCATCTGAC 1860  
 Db ||||| 1890 TGCAAGACCTGAGATTTACGCGGGGCTCCAAAAGCATTTATGCTGGCACCATCTGAC 1949  
 Qy ||||| 1861 GTGCAAGGCTGGGGGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1920  
 Db ||||| 1950 GTGCAAGGCTGGGGGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 2009  
 Qy ||||| 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGGTATGATAA 1980  
 Db ||||| 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGGTATGATAA 2069  
 Qy ||||| 1981 TACATGTGACAGCTTTCTGTTCAACTTTGAACATTAATTTTGTGTGATGACAAACCGGAG 2040  
 Db ||||| 2070 TACATGTGACAGCTTTCTGTTCAACTTTGAACATTAATTTTGTGTGATGACAAACCGGAG 2129  
 Qy ||||| 2041 GGTAAACAAATTTGTTTGGCAATCATTCGGTAAATCCAAATCTGTATGCAAAAGTTATG 2100  
 Db ||||| 2130 GGTAAACAAATTTGTTTGGCAATCATTCGGTAAATCCAAATCTGTATGCAAAAGTTATG 2189  
 Qy ||||| 2101 ATGTTTAAACGTTGATCAAGAGTATGATTTTTCCTCAAGAGAGCATTCAGACTGGGAA 2160  
 Db ||||| 2190 ATGTTTAAACGTTGATCAAGAGTATGATTTTTCCTCAAGAGAGCATTCAGACTGGGAA 2249  
 Qy ||||| 2161 GAGCTGTTTTTGTATTAAGATTAAGCAGGCTGATAGCCCTGGAAGTATGTCCGATGAA 2220  
 Db ||||| 2250 GAGCTGTTTTTGTATTAAGATTAAGCAGGCTGATAGCCCTGGAAGTATGTCCGATGAA 2309  
 Qy ||||| 2221 AGAGAAATGGAATCCCTTGA 2241  
 Db ||||| 2310 AGAGAAATGGAATCCCTTGA 2330

RESULT 5  
 HSU61145 2600 bp mRNA linear PRI 11-JUL-1997  
 LOCUS HSU61145  
 DEFINITION Human enhancer of zeste homolog 2 (EZH2) mRNA, complete cds.  
 ACCESSION U61145  
 VERSION U61145.1 GI:1575348  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 REFERENCE  
 AUTHORS Laible,G., Wolf,A., Dorn,R., Reuter,G., Nislow,C., Lebersorger,A.,  
 Popkin,D., Pillus,I. and Jenwein,T.  
 Mammalian homologues of the Polycomb-group gene Enhancer of zeste  
 mediate gene silencing in Drosophila heterochromatin and at S.  
 cerevisiae telomeres  
 JOURNAL EMBO J. 16 (11), 3219-3232 (1997)  
 PUBMED 9214638  
 REFERENCE  
 AUTHORS Laible,G., Lebersorger,A. and Jenwein,T.



TITLE Direct Submission  
JOURNAL Submitted (18-JUN-1996) I.M.P., Dr. Bohrgasse 7, Vienna A-1030,  
Austria

FEATURES  
source Location/Qualifiers

1..2600  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="BJA-B"  
/cell\_type="B-cell"  
1..2600

gene  
CDS

/gene="EZH2"  
90..2330  
/gene="EZH2"  
/note="Allele: EZH1; homolog of the Drosophila polycomb  
group protein enhancer of zeste encoded by GenBank  
Accession Number U00180"  
/codon\_start=1  
/product="enhancer of zeste homolog 2"  
/protein\_id="AAC51520.1"  
/db\_xref="GI:1575349"

/translation="MGQTGKSKGKPVCKRKYKSEYMLRQLKPRBRADYKMFSS  
NRQKLIBRIBILNOMKORLIOPVHILSVSLRGTRBCSTVSDLPFOVLPKTLN  
AVASVIMTSMPLQONFVDETVLHNIPLYGDKVLDODGTFIELINKYKVGHD  
RBCGFINDBI FVELYNALGVYDNDGDDGDDPBRBRKOKDLBDHDDKESRPRKF  
BSDKIPEALISMFPDKTRAEIKKKYKELTBQALPRECTPHIDPMKSVOREO  
SLHSFTLFCRCFCFKYDCFLPHFPAHPENTYKKTETALDNKCPGQCTOHLGAKEF  
AAALTAERIKTTPKPRGGRRRRLPNNSRSPSTPTINLVESKDTOSDRABETGEN  
NDKEBEKDETSSEANSRCOTPIKMKPNLEPPENYSGAKASMSRVILGTYDN  
FCALAIRIGTKTCROYEYEVKESIIAPADVDTPRKKRGRKMRALGRIOL  
KDDGSNNHYNYOPCDHPROPDSSCPVIAQNPCKEPCGSSSECONPPGRCRCAOC  
NTKQPCYLAARBCPDCLCTGGAADHDSKRVSKNSIQGSKKHLIAPSDVAGN  
GIFIDPVQKNFISYICGELISQDEADRGKVDYKIMCSLFNLNNDPVDAIKGN  
KIRFNHSHVNPICVAKVMWVNDHRIIGIPAKRAIQTGELFPDYRYSOADMLKVGIE  
REMRIP"

## ORIGIN

Query Match 100.0%; Score 2241; DB 5; Length 2600;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGGCCAGACTGGGAAGAAATCTGAGAGGAGCAGTTGTTGGCGGAAGCGTGTAA 60  
2 ATGGCCAGACTGGGAAGAAATCTGAGAGGAGCAGTTGTTGGCGGAAGCGTGTAA 149  
3 TCAGAGTACATCGACTGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAATGAAGT 209  
4 ATGTTAGTTCGAATCGTGAAGAAATTTGGAAAGACGGAATCTTAAACCAAGATG 180  
5 ATGTTAGTTCGAATCGTGAAGAAATTTGGAAAGACGGAATCTTAAACCAAGATG 269  
6 AAACGCGAAGGATCAGCTGTGCACTCTGAGTCTGAGGCTCATTGGCGGAGT 240  
7 AAACGCGAAGGATCAGCTGTGCACTCTGAGTCTGAGGCTCATTGGCGGAGT 329  
8 AGGAGTGTTCGGTGAACAGTGTGATTTTCCACAAGAGTATCCCATTTAAAGCT 300  
9 AGGAGTGTTCGGTGAACAGTGTGATTTTCCACAAGAGTATCCCATTTAAAGCT 389  
10 CTGAATGCACTTCACTGATACCAATATGTTCTGGTCTCCCTTACAGCAATTTT 360  
11 CTGAATGCACTTCACTGATACCAATATGTTCTGGTCTCCCTTACAGCAATTTT 449  
12 ATGTGGAAGATGAACGTTTATCATTAATCTCTATATGGAAGATGATTTTAT 420  
13 ATGTGGAAGATGAACGTTTATCATTAATCTCTATATGGAAGATGATTTTAT 509  
14 CAGATGCTATCTTCACTGATGAACATTAATTAATTAATGATGGGAAGTACACGGGAT 480  
15 CAGATGCTATCTTCACTGATGAACATTAATTAATTAATGATGGGAAGTACACGGGAT 569

481 AGAGATGTGGGTTTATAATGATGAATTTTGTGAGTTGTGATGATCCCTGTGCA 540  
482 AGAGATGTGGGTTTATAATGATGAATTTTGTGAGTTGTGATGATCCCTGTGCA 629  
541 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
542 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
601 AAAGATCTGGAGATCAACCGAGATGATGATGATGATGATGATGATGATGATGATGAT 660  
602 AAAGATCTGGAGATCAACCGAGATGATGATGATGATGATGATGATGATGATGATGAT 749  
661 GATAAATTTTGAAGCATTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 720  
662 GATAAATTTTGAAGCATTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 809  
721 AAGGAAATTAATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGATGAT 780  
722 AAGGAAATTAATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGATGAT 869  
781 ACCCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
782 ACCCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929  
841 TTGATAGCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
842 TTGATAGCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 989  
901 GCAACACCCCAACTTATTAAGCGAGAGAGACAGAAACAGCTCTAGACAAACCTTGT 960  
902 GCAACACCCCAACTTATTAAGCGAGAGAGACAGAAACAGCTCTAGACAAACCTTGT 1049  
961 GCAACACCTGTTACAGCAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
962 GCAACACCTGTTACAGCAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109  
1021 GAGCGGATTAAGACCCCAACCTCAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
1022 GAGCGGATTAAGACCCCAACCTCAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169  
1081 AACAGTACAGGCGGAG 1140  
1082 AACAGTACAGGCGGAG 1229  
1141 GATGAG 1200  
1142 GATGAG 1289  
1201 AAAGATGAACCTTGAGAGCTCTCTGAGAGCAATTTCTGGTCTCAACCAATTAAGATG 1260  
1202 AAAGATGAACCTTGAGAGCTCTCTGAGAGCAATTTCTGGTCTCAACCAATTAAGATG 1349  
1261 AAGCGAATTAATTAAGCTCTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
1262 AAGCGAATTAATTAAGCTCTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409  
1321 AAGCTCTCATTTGAGACTTATGATGAACATTTCTGTGCTCAATGCTTAAATGAGAGC 1380  
1322 AAGCTCTCATTTGAGACTTATGATGAACATTTCTGTGCTCAATGCTTAAATGAGAGC 1469  
1410 AAGCTCTCATTTGAGACTTATGATGAACATTTCTGTGCTCAATGCTTAAATGAGAGC 1440  
1381 AAACATGTATGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
1382 AAACATGTATGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1529  
1470 AAACATGTATGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1529  
1441 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
1442 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589  
1530 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589  
1501 GCAACCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1502 GCAACCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1649  
1590 GCAACCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1649  
1561 CAACCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620

DB 1650 CAACCTGTATGATCCACCGGACCTTGTGACAGTTGCTGTCCTTGTGATGACAA 1709  
QY 1621 AATTTTGTGAAAAGTTTGTGATGTAAGTTGACAGTGTCAAACCCCTTCCGGAGTC 1680  
DB 1710 AATTTTGTGAAAAGTTTGTGATGTAAGTTGACAGTGTCAAACCCCTTCCGGAGTC 1769  
QY 1681 CGCTGCAAGCAGAGTGACACCAAGAGTCCCGTGTCACTGCTGTCCGAGAGTGT 1740  
DB 1770 CGCTGCAAGCAGAGTGACACCAAGAGTCCCGTGTCACTGCTGTCCGAGAGTGT 1829  
QY 1741 GACCCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATTGGACAGTAAATGTGTCC 1800  
DB 1830 GACCCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATTGGACAGTAAATGTGTCC 1889  
QY 1801 TGCAGAACTGCAATTTACAGCGGGGCTCCAAAAGCATCTATTTGCTGGCACTTGCAC 1860  
DB 1890 TGCAGAACTGCAATTTACAGCGGGGCTCCAAAAGCATCTATTTGCTGGCACTTGCAC 1949  
QY 1861 GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGAGAAAATGAATTCATCTCAGAA 1920  
DB 1950 GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGAGAAAATGAATTCATCTCAGAA 2009  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGACCTGACAGAGAGGAAAAGTATGATGAA 1980  
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGACCTGACAGAGAGGAAAAGTATGATGAA 2069  
QY 1981 TACATGTGACACTTTCTGTTCACATTTGAACAATGATTTTGTGTGAGTGAACCGCGAAG 2040  
DB 2070 TACATGTGACACTTTCTGTTCACATTTGAACAATGATTTTGTGTGAGTGAACCGCGAAG 2129  
QY 2041 GGTAAACAATTTGTTTGAATCAATTCGCTAAATCCAACTGCTATGCAAAAGTTATG 2100  
DB 2130 GGTAAACAATTTGTTTGAATCAATTCGCTAAATCCAACTGCTATGCAAAAGTTATG 2189  
QY 2101 ATGTTAAACGATGATCAGAGATGATATTTTGTGCAAGAGAGCCATCAGACTGGCGAA 2160  
DB 2190 ATGTTAAACGATGATCAGAGATGATATTTTGTGCAAGAGAGCCATCAGACTGGCGAA 2249  
QY 2161 GAGCTGTTTTTGAATTAAGATGACGAGGCTGATGCTGAAAGTATGTGGCATGCAA 2220  
DB 2250 GAGCTGTTTTTGAATTAAGATGACGAGGCTGATGCTGAAAGTATGTGGCATGCAA 2309  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
DB 2310 AGAGAAATGAAATCCCTTGA 2330

RESULT 6  
BD186247  
LOCUS BD186247 2558 bp DNA linear PAT 17-JUN-2003  
DEFINITION STAT6 activating gene.  
ACCESSION BD186247  
VERSION BD186247.1 GI:31878447  
KEYWORDS WO 02096943-A/116.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 2558)  
Honda, G., Matsumura, A., Muramatsu, S. and Ishizawa, K.  
STAT6 activating gene  
Patent: WO 02096943-A 116 05-DEC-2002;  
ASABI KASEI CORP, GOICHI HONDA, AKIO MATSUMURA, SHUJI MURAMATSU, KENYA  
ISHIZAWA  
COMMENT OS Homo sapiens (human)  
PN WO 02096943-A/116  
PD 05-DEC-2002  
PF 22-MAY-2002 WO 2002JP004949  
PR 23-MAY-2001 JP 01P 157043.30-AUG-2001 JP 01P 260681 PR  
10-OCT-2001 JP 01P 313175  
PI GOICHI HONDA, AKIO MATSUMURA, SHUJI MURAMATSU, KENYA ISHIZAWA PC

COX14/47, COX17/00, C12N15/12, C12N5/10, C12P21/02, C12P21/08, PC  
C1205/02,  
PC A61P3/06, A61P3/10, A61P29/00, A61P31/00, A61P35/00, A61P37/00, PC  
G01N33/15,  
PC G01N33/50, G01N33/53, G01N33/56, G06F17/30, G06F17/60 CC STAT6  
activating gene  
FT CDS Location/Qualifiers  
1..2558  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/DB\_xref="taxon:9606"

ORIGIN  
Query Match 99.84; Score 2236.2; DB 2; Length 2558;  
Best Local Similarity 99.94; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGGCCGACCTGGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGAAAGCTGTAA 60  
58 ATGGCCGACCTGGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGAAAGCTGTAA 117  
61 TCGAGTACATGCGACCTGACAGAGCTCAAGAGGTTGAGAGCTGATGAAGTAAAGT 120  
118 TCGAGTACATGCGACCTGACAGAGCTCAAGAGGTTGAGAGCTGATGAAGTAAAGT 177  
121 ATGTTTGTGTCATCTGTCAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180  
178 ATGTTTGTGTCATCTGTCAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 237  
181 AAACAGGAGAGATACAGCTGTGACATCTGATCTTCTGTGAGCTATTCGCGGAGCT 240  
238 AAACAGGAGAGATACAGCTGTGACATCTGATCTTCTGTGAGCTATTCGCGGAGCT 297  
241 AGGAGGCTTCCGGGACCAAGCTTGGATTTTCCAAACAAGTCAATCCCATTAAGACT 300  
298 AGGAGGCTTCCGGGACCAAGCTTGGATTTTCCAAACAAGTCAATCCCATTAAGACT 357  
301 CTGAATCAGTGTGCTTCAATACCAATATGATTTCTGCTCCCTACAGCAAGATTTT 360  
358 CTGAATCAGTGTGCTTCAATACCAATATGATTTCTGCTCCCTACAGCAAGATTTT 417  
418 ATGTTGAGAGATGAACCTGTTTACATTAATCTTATATGAGAGATGAAGTTTATGAT 477  
421 CAGATGATCTTCAATTAAGAACTAATTAAGAAATTAATGAGGAAAGTACACGGGAT 480  
478 CAGATGATCTTCAATTAAGAACTAATTAAGAAATTAATGAGGAAAGTACACGGGAT 537  
481 AGAGATGAGGCTTATTAATGATGAATTTTGTGAGGTTGTGAATGCCCTTGTCAA 540  
538 AGAGATGAGGCTTATTAATGATGAATTTTGTGAGGTTGTGAATGCCCTTGTCAA 597  
541 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
601 AAAGATCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
658 AAAGATCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
661 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGACAGCAAGCAAGCACTA 720  
718 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGACAGCAAGCAAGCACTA 777  
721 AAAGAAATATTAAGAACTACAGCAAGCAAGCTCCAGGCGCACTTCTCTGATATGT 780  
778 AAAGAAATATTAAGAACTACAGCAAGCAAGCTCCAGGCGCACTTCTCTGATATGT 837  
781 ACCCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

Db 838 ACCCCCAATAGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 897  
 Qy 841 TTTTCAATGCTTTCTGTAGGCGAATTTTAAATATGATGCTTCTTACATCTTTTCAAT 900  
 Db 898 TTTTCAATGCTTTCTGTAGGCGAATTTTAAATATGATGCTTCTTACATCTTTTCAAT 957  
 Qy 901 GCAACACCCCAAGCTTATAGGCGAAGAACACAGAAACAGCTCTTGAACAACAACTTGT 960  
 Db 958 GCAACACCCCAAGCTTATAGGCGAAGAACACAGAAACAGCTCTTGAACAACAACTTGT 1017  
 Qy 961 GGAACAAGTGTATACAGATTTGAGGAGCAAAAGATTGTCTGCTCTCAACGCT 1020  
 Db 1018 GGAACAAGTGTATACAGATTTGAGGAGCAAAAGATTGTCTGCTCTCAACGCT 1077  
 Qy 1021 GAGCGGATTAAGACCCCAACCAAGCTTCAAGAGGCGCGAAGAGAGAGCTTCCCAAT 1080  
 Db 1078 GAGCGGATTAAGACCCCAACCAAGCTTCAAGAGGCGCGAAGAGAGAGCTTCCCAAT 1137  
 Qy 1081 AACATAGAGGCGCGACACCCCAACCAATTAATGTCTGGAATCAAAAGATACAGACAT 1140  
 Db 1138 AACATAGAGGCGCGACACCCCAACCAATTAATGTCTGGAATCAAAAGATACAGACAT 1197  
 Qy 1141 GATAGGAGACAGAGGACTGAAACGCGGCGAGAGAAACAATGATTAAGAGAGAGAGAG 1200  
 Db 1198 GATAGGAGACAGAGGACTGAAACGCGGCGAGAGAAACAATGATTAAGAGAGAGAGAG 1257  
 Qy 1201 AAAGATGAATCTTCAAGCTCTCTGTAGCAAAATCTCGGTCTCAACCAATTAAGATG 1260  
 Db 1258 AAAGATGAATCTTCAAGCTCTCTGTAGCAAAATCTCGGTCTCAACCAATTAAGATG 1317  
 Qy 1261 AAGCAATATATGAACCTCTGAGAAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 1318 AAGCAATATATGAACCTCTGAGAAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377  
 Qy 1321 AGAGTCTCTCATTTGCACTTACTATGACAAATTTCTGTGCAATGTCTAGGTTAATTTG 1380  
 Db 1378 AGAGTCTCTCATTTGCACTTACTATGACAAATTTCTGTGCAATGTCTAGGTTAATTTG 1437  
 Qy 1381 AAAACATGTAGACAGGTGTATGATTTAGAGTCAAAAGAACTTACATCATAGCTCAAGT 1440  
 Db 1438 AAAACATGTAGACAGGTGTATGATTTAGAGTCAAAAGAACTTACATCATAGCTCAAGT 1497  
 Qy 1441 CCGGCTGAGAGATGATGATCTCTCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db 1498 CCGGCTGAGAGATGATGATCTCTCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1557  
 Qy 1501 GCAACATGCAAGAAATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 Db 1558 GCAACATGCAAGAAATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 Qy 1561 CAACCTGTGATCATTCACGCGAGCTTGTGACAGTCTGTGCTTGTGTGATGACACAA 1620  
 Db 1618 CAACCTGTGATCATTCACGCGAGCTTGTGACAGTCTGTGCTTGTGTGATGACACAA 1677  
 Qy 1621 AATTTTGTGAAAGATTTTGTGCAATGTAGTTCAAGGTGTCAAAACGCTTCCGGAAGC 1680  
 Db 1678 AATTTTGTGAAAGATTTTGTGCAATGTAGTTCAAGGTGTCAAAACGCTTCCGGAAGC 1737  
 Qy 1681 CGCTGCAAGACAGTGTGACACACAGAGAGTCCGCTGTACCTGGCTGTCCGAGAGGT 1740  
 Db 1738 CGCTGCAAGACAGTGTGACACACAGAGAGTCCGCTGTACCTGGCTGTCCGAGAGGT 1797  
 Qy 1741 GACCTGACCTTCTGTCTTACTTGTGAGCCGCTGACCAATGGGACAGTAAATGTGTCC 1800  
 Db 1798 GACCTGACCTTCTGTCTTACTTGTGAGCCGCTGACCAATGGGACAGTAAATGTGTCC 1857  
 Qy 1801 TGCAAGAACTGTGCAATTCAGCGGCGCTCAAAAGAGATCTATTGCTGTGCAACATCTAC 1860  
 Db 1858 TGCAAGAACTGTGCAATTCAGCGGCGCTCAAAAGAGATCTATTGCTGTGCAACATCTAC 1917  
 Qy 1861 GTGAGAGCTGTGGGAGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1920  
 Db 1918 GTGAGAGCTGTGGGAGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1977

Qy 1921 TACTGTGAGAGATTAATTTCTAAGATGAGCTGACAGAGAGGGAAGGTGATGATAAA 1980  
 Db 1978 TACTGTGAGAGATTAATTTCTAAGATGAGCTGACAGAGAGGGAAGGTGATGATAAA 2037  
 Qy 1981 TACATGTGACAGCTTCTGTTCACATGGAACATGATTTTGTGTGTGATGCAACCCGCAAG 2040  
 Db 2038 TACATGTGACAGCTTCTGTTCACATGGAACATGATTTTGTGTGTGATGCAACCCGCAAG 2097  
 Qy 2041 GGTAAACAATTCGTTTTCGAAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2100  
 Db 2098 GGTAAACAATTCGTTTTCGAAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2157  
 Qy 2101 ATGTTTAAGGTGATCAAGATGATGATTTTGTGCAAGAGAGCATGCACTGGCGAA 2160  
 Db 2158 ATGTTTAAGGTGATCAAGATGATGATTTTGTGCAAGAGAGCATGCACTGGCGAA 2217  
 Qy 2161 GAGCTGTTTGTGATTAACAGATACAGCGAGCTGATGCTGAAATGATGCGCATCGAA 2220  
 Db 2218 GAGCTGTTTGTGATTAACAGATACAGCGAGCTGATGCTGAAATGATGCGCATCGAA 2277  
 Qy 2221 AGAGAAATGAAATCCCTTGA 2241  
 Db 2278 AGAGAAATGAAATCCCTTGA 2298

RESULT 7  
 AA821950 2576 bp DNA linear PAT 10-DEC-2003  
 LOCUS AX821950  
 DEFINITION Sequence 78 from Patent WO03068961.  
 ACCESSION AX821950  
 VERSION AX821950.1 GI:39725171  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCES  
 1 Andrews, P.A., Walsh, J.A. and Gokhale, P.A.  
 Method to modify differentiation of pluripotent stem cells  
 Patent: WO 03068961-A 78 21-AUG-2003;  
 Axordia Limited (GB)

FEATURES  
 source location/Qualifiers  
 1..2576  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 99.8%; Score 2236.2; DB 2; Length 2576;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGGGCCAGATGGGAAGAAATCTGAGAGGAGCAAGTTGTCGCGAAGCGGTAA 60  
 Db 58 ATGGGCCAGATGGGAAGAAATCTGAGAGGAGCAAGTTGTCGCGAAGCGGTAA 117  
 Qy 61 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAGACGAGCTGATGAAGTAAAGAT 120  
 Db 118 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAGACGAGCTGATGAAGTAAAGAT 177  
 Qy 121 ATGTTTAGTTCGAATCGTCAGAAAATTTGGAAGAAAGGAAATCTTAAACCAAGATG 180  
 Db 178 ATGTTTAGTTCGAATCGTCAGAAAATTTGGAAGAAAGGAAATCTTAAACCAAGATG 237  
 Qy 181 AAACAGGAGAGATACAGCTGTGACATCTGCTTCTGTGAGCTCATTTGGGAGGACT 240  
 Db 238 AAACAGGAGAGATACAGCTGTGACATCTGCTTCTGTGAGCTCATTTGGGAGGACT 297  
 Qy 241 AGGAGTGTTCGATGACAGATGATTTTTCACACAGATCATTCCTTAAAGACT 300  
 Db 298 AGGAGTGTTCGATGACAGATGATTTTTCACACAGATCATTCCTTAAAGACT 357

Qy	301	CTGAATGAGTTGGCTTGAGTACCATTAATGATCTTGATCTCCCTTACAGCAGAAATTTT	360	Qy	1381	AAAAATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGCATCATAGCTCCAGCT	1440
Db	358	CTGAATGAGTTGGCTTGAGTACCATTAATGATCTTGATCTCCCTTACAGCAGAAATTTT	417	Db	1438	AAAAATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGCATCATAGCTCCAGCT	1497
Qy	361	ATGCTGGAAGATGAACTGTTTTCATTAACATTCCTTAATATGAGATGAAAGTTTATAT	420	Qy	1441	CCCGCTGAGAGATGTGATATCTCTCCAGAGAAAAGAGGAAAACACCGGTTGTGGGCT	1500
Db	418	ATGCTGGAAGATGAACTGTTTTCATTAACATTCCTTAATATGAGATGAAAGTTTATAT	477	Db	1498	CCCGCTGAGAGATGTGATATCTCTCCAGAGAAAAGAGGAAAACACCGGTTGTGGGCT	1557
Qy	421	CAGATGTGATCTTCACTTGAAGAACTAAATTAATTAATGATGAGAAAGTACAGGGGAT	480	Qy	1501	GCACTGTCAGAAAGATACAGCTGTAAGAAAGAGAGCGGCTCTCTAACATGTTTCAACTAT	1560
Db	478	CAGATGTGATCTTCACTTGAAGAACTAAATTAATTAATGATGAGAAAGTACAGGGGAT	537	Db	1558	GCACTGTCAGAAAGATACAGCTGTAAGAAAGAGAGCGGCTCTCTAACATGTTTCAACTAT	1617
Qy	481	AGAGATGTGGGTTTATTAATGATGAAATTTTGTGAGTGTGTAATGACCTTGTCTCA	540	Qy	1561	CAACCTGTGATCATCCAGCGAGCTTGTGACAGTTGCGTCCCTGTGTGATGACAA	1620
Db	538	AGAGATGTGGGTTTATTAATGATGAAATTTTGTGAGTGTGTAATGACCTTGTCTCA	597	Db	1618	CAACCTGTGATCATCCAGCGAGCTTGTGACAGTTGCGTCCCTGTGTGATGACAA	1677
Qy	541	TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	600	Qy	1621	AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATG	1680
Db	598	TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	657	Db	1678	AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATG	1737
Qy	601	AAAGATCTGAGAGATCAACCGAGATGATTAAGAAAGCGGCGCACTCGGAAATTTCTCT	660	Qy	1681	CGCTGCAAGCAGAGTCAACCAAGCAGTGCCTGTACTACCTGCTGTCCGAGAGTGT	1740
Db	658	AAAGATCTGAGAGATCAACCGAGATGATTAAGAAAGCGGCGCACTCGGAAATTTCTCT	717	Db	1738	CGCTGCAAGCAGAGTCAACCAAGCAGTGCCTGTACTACCTGCTGTCCGAGAGTGT	1797
Qy	661	GATATAATTTTGAAGCAATTCCTCATGTTTCCAGATTAAGGCAACAGAGAAACTA	720	Qy	1741	GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGACAGTAAATGTGTCC	1800
Db	718	GATATAATTTTGAAGCAATTCCTCATGTTTCCAGATTAAGGCAACAGAGAAACTA	777	Db	1798	GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGACAGTAAATGTGTCC	1857
Qy	721	AAGAAAAATATTAAGAACTCACCGAACAGAGCTCCAGGCGCATTTCTCTCTGATGT	780	Qy	1801	TGCAAGAACTGCAAGTATTCAGCGGGGCTCCAAAAGCATATATGCTGGACCATCTGAC	1860
Db	778	AAGAAAAATATTAAGAACTCACCGAACAGAGCTCCAGGCGCATTTCTCTCTGATGT	837	Db	1858	TGCAAGAACTGCAAGTATTCAGCGGGGCTCCAAAAGCATATATGCTGGACCATCTGAC	1917
Qy	781	ACCCCAACATGATGAGCAAAATGCTAATCTGTGAGAGAGAGCAAAAGCTTCACTCC	840	Qy	1861	GTGGCAGCTGTGGGATTTTATCAAAAGATCTGTGAGAGAGAGAGAGAGATTCATCTCA	1920
Db	838	ACCCCAACATGATGAGCAAAATGCTAATCTGTGAGAGAGAGCAAAAGCTTCACTCC	897	Db	1918	GTGGCAGCTGTGGGATTTTATCAAAAGATCTGTGAGAGAGAGAGAGAGATTCATCTCA	1977
Qy	841	TTTCAATAGCTTTTCTGAGCGATGTTTAAATATGATCTGCTTCTCATCTTTTCTAT	900	Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAAGTATGATTA	1980
Db	898	TTTCAATAGCTTTTCTGAGCGATGTTTAAATATGATCTGCTTCTCATCTTTTCTAT	957	Db	1978	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAAGTATGATTA	2037
Qy	901	GCAACACCCCAACATTAATGAGGAGAAACAGAAACAGCTCTAGACAAACCTTGT	960	Qy	1981	TACATGTGAGCTTCTGCTTCACTTGAACAAATTTTGTGAGATGACACCCGAG	2040
Db	958	GCAACACCCCAACATTAATGAGGAGAAACAGAAACAGCTCTAGACAAACCTTGT	1017	Db	2038	TACATGTGAGCTTCTGCTTCACTTGAACAAATTTTGTGAGATGACACCCGAG	2097
Qy	961	GACACACGCTTTCACAGCACTTGGAGGAGCAAAAGGTTGCTGCTCTCAACGCT	1020	Qy	2041	GCTAACAAATTCCTTGTGCAATCATTCGTAATCAACTGCTATGCAAAAGTTATG	2100
Db	1018	GACACACGCTTTCACAGCACTTGGAGGAGCAAAAGGTTGCTGCTCTCAACGCT	1077	Db	2098	GCTAACAAATTCCTTGTGCAATCATTCGTAATCAACTGCTATGCAAAAGTTATG	2157
Qy	1021	GAGCGATTAAGAACCCCAACAAAGCTCAGAGAGCGCAGAGAGAGAGAGAGCTTCCCAT	1080	Qy	2101	ATGCTTAACGCTGATCAAGATAGTATTTTTCAGAGAGAGAGAGAGAGCTTCCGAG	2160
Db	1078	GAGCGATTAAGAACCCCAACAAAGCTCAGAGAGCGCAGAGAGAGAGAGAGCTTCCCAT	1137	Db	2158	ATGCTTAACGCTGATCAAGATAGTATTTTTCAGAGAGAGAGAGAGAGCTTCCGAG	2217
Qy	1081	AAAGTACAGGCGCAGACCCCAACATTAATGCTGAGATCAAGAGATACAGACGT	1140	Qy	2161	GAGCTGTTTTTGTATACAGTACAGCAGAGCTGATGCTCCGTAAGTATGCTGAGTCAA	2220
Db	1138	AAAGTACAGGCGCAGACCCCAACATTAATGCTGAGATCAAGAGATACAGACGT	1197	Db	2218	GAGCTGTTTTTGTATACAGTACAGCAGAGCTGATGCTCCGTAAGTATGCTGAGTCAA	2277
Qy	1141	GATGAGGAG	1200	Qy	2221	AGAGAAATGGAATCCCTTGA	2241
Db	1198	GATGAGGAG	1257	Db	2278	AGAGAAATGGAATCCCTTGA	2298
Qy	1201	AAAGATGAACTTCAAGCTCTCTGAGCAAAATTCCTGCTCAAAACCAATTAAGATG	1260	<div>RESULT 8</div> <div>HSZESTENH</div> <div>LOCUS</div> <div>DEFINITION</div> <div>ACCESSION</div> <div>VERSION</div> <div>KEYWORDS</div> <div>SOURCE</div> <div>ORGANISM</div> <div>2576 bp mRNA linear PRI 18-APR-2005</div> <div>H sapiens mRNA for enhancer of zeste.</div> <div>X95653.1 GI:1438063</div> <div>zeste enhancer gene.</div> <div>Homo sapiens (human)</div> <div>Homo sapiens</div> <div>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.</div>			
Db	1258	AAAGATGAACTTCAAGCTCTCTGAGCAAAATTCCTGCTCAAAACCAATTAAGATG	1317				
Qy	1261	AAAGCAATATTTGAATCTCTGAGAAATGTGAGTGTGAGTGTGCTGAAGCTCAATGTTT	1320				
Db	1318	AAAGCAATATTTGAATCTCTGAGAAATGTGAGTGTGAGTGTGCTGAAGCTCAATGTTT	1377				
Qy	1321	AGAGTCCATTTGGAGACTTACATGATCAATTTCTGTGCAATTTGATGATTAATTTGGAGC	1380				
Db	1378	AGAGTCCATTTGGAGACTTACATGATCAATTTCTGTGCAATTTGATGATTAATTTGGAGC	1437				

REFERENCE 1 (bases 1 to 2576)  
AUTHORS Chen, H., Roseler, C., and Antonarakis, S. E.  
TITLE Cloning of a human homolog of the Drosophila enhancer of zeste gene (EZH2) that maps to chromosome 21q22.2  
JOURNAL Genomics 38 (1), 30-37 (1996)  
PUBMED 8954776  
REFERENCE 2 (bases 1 to 2576)  
AUTHORS Antonarakis, S. E.  
TITLE Direct Submission  
SUBMITTED (13-FEB-1996) S. E. Antonarakis, Division of Medical Genetics, University of Geneva School of Medicine, Rue Michel-Servet 1, CH-1211 Geneva, SWITZERLAND  
COMMENT Related protein P4224.  
FEATURES  
Location/Qualifiers  
1..2576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="q22.2"  
/issue\_type="brain"  
/dev\_stage="fetal"  
1..2576  
/gene="EZH2"  
58..2298  
/gene="EZH2"  
/codon\_start=1  
/product="enhancer of zeste"  
/protein\_id="CA64955.1"  
/db\_xref="GI:1438064"  
/db\_xref="GOA:Q15910"  
/db\_xref="UniProtKB/Swiss-Prot:Q15910"  
/translation="MGQTEKSEKSPVCCKRKVKSEYMLRLQKRPRADEVKSMFSS  
NRQKILRETEILNQEMKQRIOPVHILSVSLRGTRBCSVSDLPFOVPLKTLN  
AAVAPIMTSPDLOQNFVEDETVLHNPYMGDEVLDODGFIIEELIKYGVKHGD  
RECGFINDEIFVELNALGOYNDDDDDGDDPERBERKQKLEDRHDKESRPRKF  
PSDKLEIASMSPDPKGTAEIKERYKELTEQOLPGALPRECTPNIIDGPNKSVORQ  
SLHSHTLPCRRCFYDCILPHFAHPPTTYRKATETALDNKPCPCQYOHLEAKEF  
AAALTAERIKTPPKRPGRRRLPNNSRSTPTINLVESKDTSDRAGETGDN  
NDKEBERKDETSSEBANSRCQTPIKKPNLEBPENWESGABASMPFVILGTYYDN  
FCALARLIGTKCRQVYEFVRVSSIIAPAPADVTPPRKKGRHRLMAACRKLQ  
KKDGSNNHYNYOPCDHPQPCDSSCPVLAQNFCEKFCQSSCONRPPGCRCAQC  
NTKQPCYLAVERCDPLCTCGAADHMDSKVSCNKSIOBGSXKHLILASPDYAG  
GIFPKDPVQKNEFISYCEIISQDBADRGVYDKRMCSFLPNLNDPVYATRKGN  
KIRFANHSVNPVCAKVMVNGDRHIGIPAKKAIQYGBELFVYDSQADALKYGIE  
REMERIP"  
2558  
/gene="EZH2"

polyA\_site

ORIGIN

Query Match. 99.8%; Score 2236.2; DB 5; Length 2576;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCACTTTGTCGGCGAAGCGTGTAAA 60  
2 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCACTTTGTCGGCGAAGCGTGTAAA 117  
3 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGT 120  
4 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGT 177  
5 ATGTTAGTTCGAATCGTCAGAAAATTTTGAAGAAGCGGAATCTTAAACCAAGATGG 180  
6 ATGTTAGTTCGAATCGTCAGAAAATTTTGAAGAAGCGGAATCTTAAACCAAGATGG 237  
7 AAACAGCGGAAGATACAGCTGTGCAATCTTGTGAGCTCATTTGGCGGGACT 240  
8 AAACAGCGGAAGATACAGCTGTGCAATCTTGTGAGCTCATTTGGCGGGACT 297  
9 AGGAGTGTTCGGTGCACGATGATTTTCCAAACAAGTCAATCCATTAAAGACT 300  
10 AGGAGTGTTCGGTGCACGATGATTTTCCAAACAAGTCAATCCATTAAAGACT 357

301 CTGAATGAGTGTCTTCACTACACCATTAATGATTTCTGGTCTCCCTACAGAGATTTT 360  
302 CTGAATGAGTGTCTTCACTACACCATTAATGATTTCTGGTCTCCCTACAGAGATTTT 417  
303 ATGTGGAAGATGAAGTCTTTTACATTAATCTCTTAATGAGAGATGAAGTCTTTAGAT 420  
304 ATGTGGAAGATGAAGTCTTTTACATTAATCTCTTAATGAGAGATGAAGTCTTTAGAT 477  
305 CAGATGTACTCTTCAATGAGAACTAATTAATTAATGATGAGAAAGTACACGGGAT 480  
306 CAGATGTACTCTTCAATGAGAACTAATTAATTAATGATGAGAAAGTACACGGGAT 537  
307 AGAGATGTGGTCTTAAATGAATGAATTTTGTGAGTGTGTGAATGCTTGTGTCA 540  
308 AGAGATGTGGTCTTAAATGAATGAATTTTGTGAGTGTGTGAATGCTTGTGTCA 597  
309 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
310 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
311 AAAGATCTGAGAGATCACAGATGATTAAGAAAGCCGCCACTCGGAAATTTCTCTCT 660  
312 AAAGATCTGAGAGATCACAGATGATTAAGAAAGCCGCCACTCGGAAATTTCTCTCT 717  
313 GATTAATTTTGAAGCCATTTCTCATGTTTCCAGATTAAGGACACAGCAAGAACTA 720  
314 GATTAATTTTGAAGCCATTTCTCATGTTTCCAGATTAAGGACACAGCAAGAACTA 777  
315 AAGGAAATATTAAGAACTACAGCAGCAGCTCCAGGCGCACTTCTCTGAATGT 780  
316 AAGGAAATATTAAGAACTACAGCAGCAGCTCCAGGCGCACTTCTCTGAATGT 837  
317 ACCCCCAATATGATGACCAATGCTAATCTGTTCAAGAGACCAAGCTTACATCTC 840  
318 ACCCCCAATATGATGACCAATGCTAATCTGTTCAAGAGACCAAGCTTACATCTC 897  
319 TTTCTATGCTTTTCTGTAAGCGATGTTTAAATATGATGCTGTTCTTACATCTTTTAT 900  
320 TTTCTATGCTTTTCTGTAAGCGATGTTTAAATATGATGCTGTTCTTACATCTTTTAT 957  
321 GCAACCCCACTTATTAAGGAGAAACAGAAACAGCTCTGACCAACCAACTTGT 960  
322 GCAACCCCACTTATTAAGGAGAAACAGAAACAGCTCTGACCAACCAACTTGT 1017  
323 GCAACCCCACTTATTAAGGAGAAACAGAAACAGCTCTGACCAACCAACTTGT 1020  
324 GCAACCCCACTTATTAAGGAGAAACAGAAACAGCTCTGACCAACCAACTTGT 1077  
325 GCAACCCCACTTATTAAGGAGAAACAGAAACAGCTCTGACCAACCAACTTGT 1080  
326 GATGAGGAG 1137  
327 GATGAGGAG 1140  
328 GATGAGGAG 1197  
329 GATGAGGAG 1257  
330 GATGAGGAG 1260  
331 GATGAGGAG 1317  
332 GATGAGGAG 1320  
333 GATGAGGAG 1377  
334 GATGAGGAG 1380  
335 GATGAGGAG 1437

Qy	1381	AAAACATGTACACAGGCTATGAGTTTGAAGTCAAAAGAACTAGATCATAGTCCAGCT	1440.0		
Db	1438	AAAACATGTACACAGGCTATGAGTTTGAAGTCAAAAGAACTAGATCATAGTCCAGCT	1497		
Qy	1441	CCCGCTGAGATGTGTGATCTCTTCCAAGAAAAAGAGAAAAACACGGTTGTGGCT	1500		
Db	1498	CCCGCTGAGATGTGTGATCTCTTCCAAGAAAAAGAGAAAAACACGGTTGTGGCT	1557		
Qy	1501	GCAACTGTCAAAAAGATACAGCTGAAAAAGACGGCTCTCTTAAACATGTTTACAATAT	1566		
Db	1558	GCACACTGTCAAAAAGATACAGCTGAAAAAGACGGCTCTCTTAAACATGTTTACAATAT	1617		
Qy	1561	CAACCCGTGTATTCACACGGCAGCCTGTGTACAGTTGTGTCCCTTGTGTATGTACAA	1620		
Db	1618	CAACCCGTGTATTCACACGGCAGCCTGTGTACAGTTGTGTCCCTTGTGTATGTACAA	1677		
Qy	1621	AATTTTGTGAAAAAGTTTGTCAATGTAGTTCAAGATGTCAAAAACGGCTTCCGGAGTGC	1680		
Db	1678	AATTTTGTGAAAAAGTTTGTCAATGTAGTTCAAGATGTCAAAAACGGCTTCCGGAGTGC	1737		
Qy	1681	CGCTGTCAAGAACAAGTGTACACCAAGACGTGCCGTGTACTTGGCTGTCCAGAGATGT	1740		
Db	1738	CGCTGTCAAGAACAAGTGTACACCAAGACGTGCCGTGTACTTGGCTGTCCAGAGATGT	1797		
Qy	1741	GACCTGTACCTCTGTCTTACTTTGTGTGAGACCCGTGACCAATTGGGACAGTAAAAATGTGTCC	1800		
Db	1798	GACCTGTACCTCTGTCTTACTTTGTGTGAGACCCGTGACCAATTGGGACAGTAAAAATGTGTCC	1857		
Qy	1801	TGCAAGAAGCTGTAGTATTCAGCGGGGGCTCCAAAAAGCATCTATTGTCTGGCCACATCTGAC	1866		
Db	1858	TGCAAGAAGCTGTAGTATTCAGCGGGGGCTCCAAAAAGCATCTATTGTCTGGCCACATCTGAC	1917		
Qy	1861	GTGGCAGGCTGGGGGATTTTATCAAAAGATCCTGTGCAAAAAATGAATTCATCTCAGAA	1920		
Db	1918	GTGGCAGGCTGGGGGATTTTATCAAAAGATCCTGTGCAAAAAATGAATTCATCTCAGAA	1977		
Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGTGTACAGAAAGGAAAAAGTATATGATTAA	1980		
Db	1978	TACTGTGAGAGATTAATTTCTCAAGATGAAGTGTACAGAAAGGAAAAAGTATATGATTAA	2037		
Qy	1981	TACATGTGACAGCTTCTGTTCACAATCTTGAACATGATTTTGTGTGTGATGTCAACCCGCAAG	2040		
Db	2038	TACATGTGACAGCTTCTGTTCACAATCTTGAACATGATTTTGTGTGTGATGTCAACCCGCAAG	2097		
Qy	2041	GCTAACAAATTCGTTTGTGCAAAATCATTCGGTAAATCCAAACTGTGTATGTCAAAAGTTATG	2100		
Db	2098	GCTAACAAATTCGTTTGTGCAAAATCATTCGGTAAATCCAAACTGTGTATGTCAAAAGTTATG	2157		
Qy	2101	ATGTGTAAACGGTATCACAGATATGATATTTTGTGCAAGAACCCATCTCAACTGTGGCAA	2160		
Db	2158	ATGTGTAAACGGTATCACAGATATGATATTTTGTGCAAGAACCCATCTCAACTGTGGCAA	2217		
Qy	2161	GAGCTGTTTTTGTATTACAGATATACAGACGACGCTGATGTGCCCTGAAGTATGTGGCATGTGAA	2220		
Db	2218	GAGCTGTTTTTGTATTACAGATATACAGACGACGCTGATGTGCCCTGAAGTATGTGGCATGTGAA	2277		
Qy	2221	AGAGAAATGAAATCCCTTGA 2241			
Db	2278	AGAGAAATGAAATCCCTTGA 2258			
RESULT 9					
LOCUS	DD208658	2512 bp	DNA		
DEFINITION	Expression Profile of Prostate Cancer.				
ACCESSION	DD208658	GI:85649860			
VERSION	DD208658.1	JP 200518522-A/95.			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	ORIGIN
1	(bases 1 to 2512)	Rabin, M.A., Chinnaiyan, A.M., and Surikumar, A.	Expression Profile of Prostate Cancer Patent: JP 200518522-A 95 23 JUN-2005; THE REGENTS OF THE UNIVERSITY OF MICHIGAN OS Homo sapiens PN JP 200518522-A/95	PD 23-JUN-2005	PF 02-AUG-2002 JP 2003517245	PR 02-AUG-2001 US 60/309581, 15-NOV-2001 US 60/334468, PR 01-AUG-2002 US 10/210120
					PI mark a rabin, aruru m chinnaiyan, arun surikumar CC	
					FH Key Location/Qualifiers	
					1. .2512	
					/organism="Homo sapiens"	
					/mol_type="unassigned DNA"	
					/db_xref="taxon:9606"	
Query Match	99.5%;	Score 2229;	DB 2;	Length 2512;		
Best Local Similarity	99.9%;	Pred. No. 0;				
Matches 2240;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;		
QY	1	ATGGGCCAGATCGTGGAAAGAAATCTGAGAAGGACCAAGTTGTGTGGCGGAGCGGTATAA	60			
DB	10	ATGGGCCAGATCGTGGAAAGAAATCTGAGAAGGACCAAGTTGTGTGGCGGAGCGGTATAA	69			
QY	61	TCAGAGTACATCGCAGCTGAGACAGCTCAGAGGTTCCAGACGAGCTGATGAATGAAGAGT	120			
DB	70	TCAGAGTACATCGCAGCTGAGACAGCTCAGAGGTTCCAGACGAGCTGATGAATGAAGAGT	128			
QY	121	ATGTTTAGTCCAAATCGTCAGAAAAATTTGGAAAGACGAAATCTTAAACCAAGAAATGG	180			
DB	129	ATGTTTAGTCCAAATCGTCAGAAAAATTTGGAAAGACGAAATCTTAAACCAAGAAATGG	188			
QY	181	AAACAGCGAAGAAATACAGCTGTCACATCTGCACTTCTGAGCTCATTTGGCGGGAGCT	240			
DB	189	AAACAGCGAAGAAATACAGCTGTCACATCTGCACTTCTGAGCTCATTTGGCGGGAGCT	248			
QY	241	AGGAGTGTGGGTGACCAAGTGAATTTCCACACAAATCATCCATTAAAGACT	300			
DB	249	AGGAGTGTGGGTGACCAAGTGAATTTCCACACAAATCATCCATTAAAGACT	308			
QY	301	CTGAATGCAATGCTTCAGTACCAATTAATGTAATCTTGGTCTCCCTTACAGCAAAATTTT	360			
DB	309	CTGAATGCAATGCTTCAGTACCAATTAATGTAATCTTGGTCTCCCTTACAGCAAAATTTT	368			
QY	361	ATGCTGGAAGATGAACCTGTTTACATPACATTCCTTATATGGGAGATGAAGTTTAAAT	420			
DB	369	ATGCTGGAAGATGAACCTGTTTACATPACATTCCTTATATGGGAGATGAAGTTTAAAT	428			
QY	421	CAGATGGTACTTTCATTTGAAGAACTAATAAAAAATTGATGGGAAAGTACACGGGAT	480			
DB	429	CAGATGGTACTTTCATTTGAAGAACTAATAAAAAATTGATGGGAAAGTACACGGGAT	488			
QY	481	AGAGATGTGGGTTTAAATGAATGAATTTTGTGAGTTGTTGTAATGCCCTTGGTCA	540			
DB	489	AGAGATGTGGGTTTAAATGAATGAATTTTGTGAGTTGTTGTAATGCCCTTGGTCA	548			
QY	541	TATATATGATGATGACATGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAAGCAG	600			
DB	549	TATATATGATGATGACATGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAAGCAG	608			
QY	601	AAAGATCGGAGGATCACCGAGATGATTAAGAAAGCGCCCACTCGGAAATTTCTCTCT	660			
DB	609	AAAGATCGGAGGATCACCGAGATGATTAAGAAAGCGCCCACTCGGAAATTTCTCTCT	668			
QY	661	GATTAATATTTTGAAGCCATTTCTCATATGTTTCCAGATTAAGGCGACACAGAAAGACTA	720			
DB	669	GATTAATATTTTGAAGCCATTTCTCATATGTTTCCAGATTAAGGCGACACAGAAAGACTA	728			
QY	721	AAGGAAAAATATTAAGAACTCACCGAAGCAGACTCCAGGCGCACTTCTCTGTAATGT	780			



```

Db      ||||| 729 AAGAAAAATATATAAACAACCGAAGAGAGCTCCAGGGGCACTTCTCTGTAATGT 788
Qy      ||||| 781 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840
Db      ||||| 789 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 848
Qy      ||||| 841 TTTCAATAGCTTTTCTGTAGGGAGATGTTTAAATATGATGCTGCTTCTATCATCTTTGAT 900
Db      ||||| 849 TTTCAATAGCTTTTCTGTAGGGAGATGTTTAAATATGATGCTGCTTCTATCATCTTTGAT 908
Qy      ||||| 901 GCAACACCCACACATTAATAGCGAGAAACACAGAAACAGCTCTAGACAAACAACTTGT 960
Db      ||||| 909 GCAACACCCACACATTAATAGCGAGAAACACAGAAACAGCTCTAGACAAACAACTTGT 968
Qy      ||||| 961 GGAACAAGCTTTTACAGCAATTTGAGGGAGCAAAAGAGTTTGTCTGCTTCAACCGCT 1020
Db      ||||| 969 GGAACAAGCTTTTACAGCAATTTGAGGGAGCAAAAGAGTTTGTCTGCTTCAACCGCT 1028
Qy      ||||| 1021 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGGCGGAGAAAGAGAGGCTTCCAT 1080
Db      ||||| 1029 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGGCGGAGAAAGAGAGGCTTCCAT 1088
Qy      ||||| 1081 AACAGTAGAGGCGCCGACACCCCAACATTAATGCTGCTGAATCAAGAGATACAGAGT 1140
Db      ||||| 1089 AACAGTAGAGGCGCCGACACCCCAACATTAATGCTGCTGAATCAAGAGATACAGAGT 1148
Qy      ||||| 1141 GATAGGAAAGCAAGGAGCTGAAACGGGGGAGAGAAACATGATTAAGAAAGAGAGAG 1200
Db      ||||| 1149 GATAGGAAAGCAAGGAGCTGAAACGGGGGAGAGAAACATGATTAAGAAAGAGAGAG 1208
Qy      ||||| 1201 AAGAGTAGAACTTCCAGCTCTCTGAGAAATTTCTCGGTGCAACACCATTAAGATG 1260
Db      ||||| 1209 AAGAGTAGAACTTCCAGCTCTCTGAGAAATTTCTCGGTGCAACACCATTAAGATG 1268
Qy      ||||| 1261 AAGCAAAATATTGAACCTCTGAGAGATGAGTAGAGTGAGTGAAGCTCAATGTTT 1320
Db      ||||| 1269 AAGCAAAATATTGAACCTCTGAGAGATGAGTAGAGTGAGTGAAGCTCAATGTTT 1328
Qy      ||||| 1321 AAGAGTCTCATTTGGAACCTTACTATGACAAATTTCTGTCCATTTGAGTTAATGGAGC 1380
Db      ||||| 1329 AAGAGTCTCATTTGGAACCTTACTATGACAAATTTCTGTCCATTTGAGTTAATGGAGC 1388
Qy      ||||| 1381 AAAAATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGACATCACTACCTCAGT 1440
Db      ||||| 1389 AAAAATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGACATCACTACCTCAGT 1448
Qy      ||||| 1441 CCGGCTGAGATGTGATCTCTCCCAAGAGAAAGAGAGAGAAACACCGGTTGTGGCT 1500
Db      ||||| 1449 CCGGCTGAGATGTGATCTCTCCCAAGAGAAAGAGAGAGAAACACCGGTTGTGGCT 1508
Qy      ||||| 1501 GCAACCTGCAAGAAATACAGCTGAAAGAGAGCGGCTCTTACCATGTTTACAACTAT 1560
Db      ||||| 1509 GCAACCTGCAAGAAATACAGCTGAAAGAGAGCGGCTCTTACCATGTTTACAACTAT 1568
Qy      ||||| 1561 CAACCTGTGATCATTCACGGGAGCTTGTGACAGTTGTGCTTGTGTATGACAA 1620
Db      ||||| 1569 CAACCTGTGATCATTCACGGGAGCTTGTGACAGTTGTGCTTGTGTATGACAA 1628
Qy      ||||| 1621 AATTTTGTGAAAGTTTTGTCAATGTATGTTCAAGAGTCAAAACCGCTTCCGGAGTGC 1680
Db      ||||| 1629 AATTTTGTGAAAGTTTTGTCAATGTATGTTCAAGAGTCAAAACCGCTTCCGGAGTGC 1688
Qy      ||||| 1681 CGCTGCAAGCAAGTGTGACCAAGAGAGTCCCGGTGTAACCTGCTGCCAGAGTGT 1740
Db      ||||| 1689 CGCTGCAAGCAAGTGTGACCAAGAGAGTCCCGGTGTAACCTGCTGCCAGAGTGT 1748
Qy      ||||| 1741 GACCTGACCTTGTCTTACTTGTGAGAGCGCTGACCAATGGGACAGTAAATGTGTCC 1800
Db      ||||| 1749 GACCTGACCTTGTCTTACTTGTGAGAGCGCTGACCAATGGGACAGTAAATGTGTCC 1808
Qy      ||||| 1801 TGCAAGAACTGCAATTCAGCGGGGCTCAAAAAGCATCTATTGCTGGCAACATCTAC 1860

```

```

Db      ||||| 1809 TGCAAGAACTGCAATTCAGCGGGGCTCAAAAAGCATCTATTGCTGGCAACATCTGAC 1868
Qy      ||||| 1861 GTGGCAGGCTGGGGGATTTTATCAAGATCTCTGTGACAGAAATTAATTCATCTCAGAA 1920
Db      ||||| 1869 GTGGCAGGCTGGGGGATTTTATCAAGATCTCTGTGACAGAAATTAATTCATCTCAGAA 1928
Qy      ||||| 1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAGAAAGTATGATAA 1980
Db      ||||| 1929 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAGAAAGTATGATAA 1988
Qy      ||||| 1981 TACATGTGACGCTTTCTGTTCACCTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2040
Db      ||||| 1989 TACATGTGACGCTTTCTGTTCACCTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2048
Qy      ||||| 2041 GGTACCAAAATTCGTTTGGCAATTCATTCGGTAAATCCAAATGCTATGCAAAAGTATG 2100
Db      ||||| 2049 GGTACCAAAATTCGTTTGGCAATTCATTCGGTAAATCCAAATGCTATGCAAAAGTATG 2108
Qy      ||||| 2101 ATGTTAACGGGTATCACAGATAGATTAATTTTGGCAAGAGCCATCGACTGGCGAA 2160
Db      ||||| 2109 ATGTTAACGGGTATCACAGATAGATTAATTTTGGCAAGAGCCATCGACTGGCGAA 2168
Qy      ||||| 2161 GAGCTGTTTTTGTATTAACATGATACAGCCAGGCTGATGCGCTGAAGTATGTCGCAATGAA 2220
Db      ||||| 2169 GAGCTGTTTTTGTATTAACATGATACAGCCAGGCTGATGCGCTGAAGTATGTCGCAATGAA 2228
Qy      ||||| 2221 AGAGAAATGGAATTCCTTGA 2241
Db      ||||| 2229 AGAGAAATGGAATTCCTTGA 2249

RESULT 10
BC010858
LOCUS      BC010858
DEFINITION Homo sapiens enhancer of zeste homolog 2 (Prosophi1a), transcript
            variant 1, mRNA (cDNA clone MGC:9169 IMGB:3901250), complete cds.
ACCESSION BC010858
VERSION    BC010858.2 GI:34194096
KEYWORDS   MGC.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1 (bases 1 to 2655)
            Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
            Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
            Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
            Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
            Diatchenko, L., Marzella, K., Farmer, A.A., Rubin, G.M., Hong, L.,
            Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
            Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshibayashi, S.,
            Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
            Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
            McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
            Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Holly, S.W.,
            Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
            Fahey, J., Helton, E., Kettelman, M., Madan, A.C., Rodriguez, S.,
            Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherbenko, Y.,
            Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
            Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
            Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalins, D.E.,
            Mammalian Gene Collection Program Team
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
            1247932
            2 (bases 1 to 2655)
            NIH MGC Project
            Direct Submission
            Submitted (12-JUL-2001) National Institutes of Health, Mammalian

```





QY	1246	ACACCAATPAAAGATGATAGCAAAATATTGAACTCTGTGAAGATGTGGAATGAGTGTCT	13050
Db	1383	ACACCAATPAAAGATGATAGCAAAATATTGAACTCTGTGAAGATGTGGAATGAGTGTCT	14424
QY	1306	GAAGCCTCAATGTTTAGTGAGTCTCATTTGGCACTTAATAATGACAAATTTCTGTGCCATTTGCT	13658
Db	1443	GAAGCCTCAATGTTTAGTGAGTCTCATTTGGCACTTAATAATGACAAATTTCTGTGCCATTTGCT	15020
QY	1366	AGGTTAATTGGGACCAAAAATATGTGACAGGGTGTATGAGTTTAAAGTCAAAAGATCTAGC	14235
Db	1503	AGGTTAATTGGGACCAAAAATATGTGACAGGGTGTATGAGTTTAAAGTCAAAAGATCTAGC	15628
QY	1426	ATCATAGCTCCAGCTCCCGCTGAGGATGTGGATATCTCTCCAGGAAAAAGAGAGGAAA	14858
Db	1553	ATCATAGCTCCAGCTCCCGCTGAGGATGTGGATATCTCTCCAGGAAAAAGAGAGGAAA	16222
QY	1486	CACCGGTTGTGGGCTGCACTGCAAGAAAGATACAGCTGAAAAAGAGACGGCTCTCTTAAC	15454
Db	1623	CACCGGTTGTGGGCTGCACTGCAAGAAAGATACAGCTGAAAAAGAGACGGCTCTCTTAAC	16822
QY	1546	CATGTTTCAACTATCAACCTGTGTATCATCAAGGACGCTTGTGACAGTTTCTGTGCTT	16050
Db	1663	CATGTTTCAACTATCAACCTGTGTATCATCAAGGACGCTTGTGACAGTTTCTGTGCTT	17424
QY	1606	TGTGTGATAGACAAAAATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGGTGTCAAAAC	16655
Db	1743	TGTGTGATAGACAAAAATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGGTGTCAAAAC	18020
QY	1666	CGCTTTCGGGATGCGCTGTGCAAGACAGTGTCAACACCAAGACAGTGTCCGTGTACTGT	17255
Db	1803	CGCTTTCGGGATGCGCTGTGCAAGACAGTGTCAACACCAAGACAGTGTCCGTGTACTGT	18622
QY	1726	GCTGTCCGAGAGTGTGACCTGTGACCTCTGTCTTAATGTGAGACCGCTGTGACATTTGGGAC	17858
Db	1863	GCTGTCCGAGAGTGTGACCTGTGACCTCTGTCTTAATGTGAGACCGCTGTGACATTTGGGAC	19222
QY	1786	AGTAAAAATGTGTCTCTGTGCAAGAACTGTGACAGTATTCAGCCGGGCTCCAAAAACATCTAATG	18454
Db	1923	AGTAAAAATGTGTCTCTGTGCAAGAACTGTGACAGTATTCAGCCGGGCTCCAAAAACATCTAATG	19822
QY	1846	CTGGCAACATCTGACGTGTGACAGGCTGTGGGGATTTTATCAAAAGATCTGTGTGCAAAAAAT	19050
Db	1983	CTGGCAACATCTGACGTGTGACAGGCTGTGGGGATTTTATCAAAAGATCTGTGTGCAAAAAAT	20424
QY	1966	GAATTCATCTCAGAAATACTGTGTGAGAGATTATTTCTCAAGATGGAAGCTGACAGAAAGGG	19655
Db	2043	GAATTCATCTCAGAAATACGTGTGTGAGAGATTATTTCTCAAGATGGAAGCTGACAGAAAGGG	21020
QY	1966	AAAGGTATGTAAATATACATGTGTGACGCTTCTGTCAACTTGAACAATGATTTGTGGTG	20252
Db	2103	AAAGGTATGTAAATATACATGTGTGACGCTTCTGTCAACTTGAACAATGATTTGTGGTG	21622
QY	2026	GATGCAACCCGCAAGGGTAAACAAAATTCGTTTGGCAATCAATGTGGTAAATCCAACTGC	20858
Db	2163	GATGCAACCCGCAAGGGTAAACAAAATTCGTTTGGCAATCAATGTGGTAAATCCAACTGC	22222
QY	2086	TATGCAAAAGTTATGATGTGTTAATCGGTGTATCAAGATAGGATTTTTTGTCCAAAGAGCC	21454
Db	2223	TATGCAAAAGTTATGATGTGTTAATCGGTGTATCAAGATAGGATTTTTTGTCCAAAGAGCC	22822
QY	2146	ATCCAGACTGTGCGAAGAGCTGTTTTGTGATTAACAGATACAGCCAGGCTGATGCTCCTGAG	22050
Db	2283	ATCCAGACTGTGCGAAGAGCTGTTTTGTGATTAACAGATACAGCCAGGCTGATGCTCCTGAG	23424
QY	2206	TATGTGGCATCGAAAGAGAAATGGAATCCCTTGA	2241
Db	2343	TATGTGGCATCGAAAGAGAAATGGAATCCCTTGA	2378

RESULT 11	
ABI79385	
LOCUS	
ABI79385	2619 bp
	mRNA
	linear
	PRI 17-SEP-2005

DEFINITION	Macaqa fascicularis testis cDNA clone: QcFA-18821, similar to human enhancer of zeste homolog 2 (Drosophila) (EZH2), transcript variant 2. mRNA, RefSeq: NM_152936.1.
ACCESSION	AB179385
VERSION	AB179385.1 GI:67972187
KEYWORDS	oligo capping; fig (full insert sequence).
SOURCE	Macaqa fascicularis (crab-eating macaque)
ORGANISM	Macaqa fascicularis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
REFERENCE	1 Ooada,N., Hirata,M., Tanuma,R., Kusuda,J., Hida,M., Suzuki,Y., Sugano,S., Gojobori,T., Shen,C.-K.J., Wu,C.I. and Hashimoto,K.: Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis Between Human and Cynomolgus Monkey CDNAs Mol. Biol. Evol. 22 (10), 1976-1982 (2005) 2 15944441
JOURNAL TITLE	International consortium for macaque cDNA sequencing and analysis DNA sequences of macaque genes expressed in brain or testis and its evolutionary implications unpublished
REFERENCE	3 (bases 1 to 2619) Hashimoto,K., Kusuda,J. and Sugano,S. Direct Submission Submitted (10-MAY-2004) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo, 163-8640, Japan (e-mail:khashim@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181) The international consortium for macaque cDNA sequencing and analysis consists of:Department for Virology and Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; Division of Genetic Resources, National Institute of Infectious Diseases of Japan, Tokyo, Japan; National Health Research Institute, Taipei, Taiwan; Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan; Department of Ecology & Evolution
COMMENT	University of Chicago, Chicago, IL, USA, Center for Information Biology, National Institute of Genetics of Japan, Mishima, Japan. Clone distribution: clone distribution information can be found at: <a href="http://www.nih.go.jp/yoken/genebank/">http://www.nih.go.jp/yoken/genebank/</a> Lab host: TOPI0 Vector: pMEBIS-FL3 (Acc.No. AB009864) R. Site1: DraIII (CACTGTCG) R. Site2: DraII (CACCATGG) Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGGCCCTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb.The SfiI-digested PCR product was cloned into distinct DraIII sites of pMEBIS-FL3. XhoI sites just outside the DraII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method. Libraries were made from:

OCCR: cerebellum cortex

QnpA: parietal lobe

QrrA: temporal lobe right

QELA: frontal lobe left

QmoA: medulla oblongata

QBSA: brain stem

Qora: occipital lobe right

Qrsa: testis

Custom primers were used for 5' and 3'-end sequencing. The full-insert sequencing was done by primer-walking method using ABI DNA sequencer.

Location/Qualifiers

1..2619

/organism="Macaqa fascicularis"

/mol\_type="mRNA"



Db 1773 CGCTCAAGACAGTGCACCAAGCAGTCCCGCTGCTACCTGGCCGCTCCGAGAGTGT 1832  
QY 1741 GACCTGACCTTGTCTTACTTGTGAGAGCCGCTGACCATTTGGGACAGTAAATGTGTCC 1800  
Db 1833 GACCTGACCTTGTCTTACTTGTGAGAGCCGCTGACCATTTGGGACAGTAAATGTGTCC 1892  
QY 1801 TGCAGAACTGCAATTCAGCGGGGCTCCAAAAGCATCTATTTGCTGGACCATCTGAC 1860  
Db 1893 TGCAGAACTGCAATTCAGCGGGGCTCCAAAAGCATCTATTTGCTGGACCATCTGAC 1952  
QY 1861 GTGGAGGCTGGGGGATTTTATCAAGATCTGTGAGAAAAATGAATTCATCTCAGAA 1920  
Db 1953 GTGGAGGCTGGGGGATTTTATCAAGATCTGTGAGAAAAATGAATTCATCTCAGAA 2012  
QY 1921 TACTGTGAGAGATTTATTTCTCAAGATGAAGCTGACAGAAAGGGAAAGTGTATATAA 1980  
Db 2013 TACTGTGAGAGATTTATTTCTCAAGATGAAGCTGACAGAAAGGGAAAGTGTATATAA 2072  
QY 1981 TACATGTGAGGCTTCTGTTCATCTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2040  
Db 2073 TACATGTGAGGCTTCTGTTCATCTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2132  
QY 2041 GGTAAACAAATTCGTTTTCGAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2100  
Db 2133 GGTAAACAAATTCGTTTTCGAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2192  
QY 2101 ATGTGTAACGGTATCAAGATAGTATTTTTCGCAAGAGCCATCCAGCTGGCGAA 2160  
Db 2193 ATGTGTAACGGGATCAAGATAGTATTTTTCGCAAGAGCCATCCAGCTGGCGAA 2252  
QY 2161 GAGCTGTTTTGATTTAGATGATGAGCAGCAGGCTGATGCGCTGAATGATGTGGCATGAA 2220  
Db 2253 GAGCTGTTTTGATTTAGATGATGAGCAGCAGGCTGATGCGCTGAATGATGTGGCATGAA 2312  
QY 2221 AGAGAAATGGAATTCCTTGA 2241  
Db 2313 AGAGAAATGGAATTCCTTGA 2333

RESULT 12  
AR718692 2476 bp DNA linear PAT 07-OCT-2005  
LOCUS AR718692 Sequence 1192 from patent US 6943241.  
DEFINITION AR718692  
ACCESSION AR718692.1 GI:77367646  
VERSION AR718692.1 GI:77367646  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2476)  
AUTHORS Iwagaki,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.-I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,I.K. and Maruho,Y.  
TITLE Full-length cDNA  
JOURNAL Patent: US 6943241-A 1192 13-SEP-2005;  
Research Association for Biotechnology; Tokyo;  
WO; Location/Qualifiers  
FEATURES  
source 1..2476  
/organism="Unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 89.0%; Score 1993.8; DB 2; Length 2476;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 2; Indels 117; Gaps 1;  
QY 1 ATGGGCGCAGACTGGGAGAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAGCGTGTAA 60  
Db 93 ATGGGCGCAGACTGGGAGAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAGCGTGTAA 152  
QY 61 TCAGAGTACATGCGACTGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGAT 120

Db 153 TCAGAGTACATGCGACTGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAATAGT 212  
QY 121 ATGTTAGTCCATGCTGAGAAATTTTGGAAAGACGAAATTTTAAACCAAGATGG 180  
Db 213 ATGTTAGTCCATGCTGAGAAATTTTGGAAAGACGAAATTTTAAACCAAGATGG 272  
QY 181 AAACAGGAGAGATGACGCTGTGACATCTGCTGTGAGCTCATTTGGCGGGAGT 240  
Db 273 AAACAGGAGAGATGACGCTGTGACATCTGCTGTGAGCTCATTTGGCGGGAGT 332  
QY 241 AGGAGTGTGGTGCAGCAGTGAATTTGATTTCCAAACAAATCATCCATTAAGACT 300  
Db 333 AGGA----- 337  
QY 301 CTGAATGAGTTGCTTCAAGTACCAATTAATGATTTCTGTGCTCCCTTACAGCAATTTT 360  
Db 338 ----- 337  
QY 361 ATGTGGAAGATGAACGTTTTCATTAACATTCCTTATATGGAGATGAAGTTTAGT 420  
Db 338 --GGTGAAGATGAACGTTTTCATTAACATTCCTTATATGGAGATGAAGTTTAGT 395  
QY 421 CAGATGTGACTTTTCAATTGAAGAACTAATAAAATTTATGATGGAAAGTACACGGGAT 480  
Db 396 CAGATGTGACTTTTCAATTGAAGAACTAATAAAATTTATGATGGAAAGTACACGGGAT 455  
QY 481 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGTGTGATGCTGCTGTGCA 540  
Db 456 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGTGTGATGCTGCTGTGCA 515  
QY 541 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 516 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575  
QY 601 AAAGTCTGAGAGATCAACGAGATGATTAAGAAAGCCGCACTCGGAAATTTCTTCT 660  
Db 576 AAAGTCTGAGAGATCAACGAGATGATTAAGAAAGCCGCACTCGGAAATTTCTTCT 635  
QY 661 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGTAAAGGCAACGAGAGACTA 720  
Db 636 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGTAAAGGCAACGAGAGACTA 695  
QY 721 AAAGAAAAATTAAGAACTGACGGAACGAGAGCTCCAGGCGACTTCTCTGAAATGT 780  
Db 696 AAAGAAAAATTAAGAACTGACGGAACGAGAGCTCCAGGCGACTTCTCTGAAATGT 755  
QY 781 ACCCCCAACATGATGAGCAAAATGCTTAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840  
Db 756 ACCCCCAACATGATGAGCAAAATGCTTAATCTGTTCAAGAGAGCAAAAGCTTACATCC 815  
QY 841 TTTCAATGAGCTTTTCTGAGGCGATGTTTAAATATGATGATGATGATGATGATGATGAT 900  
Db 816 TTTCAATGAGCTTTTCTGAGGCGATGTTTAAATATGATGATGATGATGATGATGATGAT 875  
QY 901 GCAACACCCCAACATTAATTAAGCGGAAGAACCAAAACAGCTTGAAGAAAGAAAGCTTGT 960  
Db 876 GCAACACCCCAACATTAATTAAGCGGAAGAACCAAAACAGCTTGAAGAAAGAAAGCTTGT 935  
QY 961 GGACCAAGTGTATACAGACTTTTGAAGGAGCAAAAGAGTTTGTGCTGCTCAGCGCT 1020  
Db 936 GGACCAAGTGTATACAGACTTTTGAAGGAGCAAAAGAGTTTGTGCTGCTCAGCGCT 995  
QY 1021 GAGCGGATTAAGACCCCAACAAAGCTTCAGAGAGCGCGAGAGAGAGAGCGCTTCCAT 1080  
Db 996 GAGCGGATTAAGACCCCAACAAAGCTTCAGAGAGCGCGAGAGAGAGAGCGCTTCCAT 1055  
QY 1081 AAACGTAGCAGGCGCCAGAGACCCCAACATTAATGCTGAGATCAAAAGATACAGACGT 1140  
Db 1056 AAACGTAGCAGGCGCCAGAGACCCCAACATTAATGCTGAGATCAAAAGATACAGACGT 1115  
QY 1141 GATAGGGAAGCAGGAGCTGAAACGGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1200

Db 1116 GATAGGAGGAGGAGCTGAAACGGGGGAGAGAACATGATTAAGAGAGAGAG 1175  
QY 1201 AAAATGAACTTGGAGCTCTGAAAGCAAAATCTGGGTGCAAAACCAATTAAGATG 1260  
Db 1176 AAAATGAACTTGGAGCTCTGAAAGCAAAATCTGGGTGCAAAACCAATTAAGATG 1235  
QY 1261 AAGCAATATATTGAACCTCTGAGAAATGAGAGTGGTGTGAAGCTCAATGTTT 1320  
Db 1236 AAGCAATATATTGAACCTCTGAGAAATGAGAGTGGTGTGAAGCTCAATGTTT 1295  
QY 1321 AGAGTCTCATTTGGCACTTACTATAGCAAAATTTCTGTGCAATTCCTAGGTTAAATGGAGCC 1380  
Db 1296 AGAGTCTCATTTGGCACTTACTATAGCAAAATTTCTGTGCAATTCCTAGGTTAAATGGAGCC 1355  
QY 1381 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATCATAGCTCCAGCT 1440  
Db 1356 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATCATAGCTCCAGCT 1415  
QY 1441 CCCGCTGAGAGTGTGATTAATCTCTTCAAGAGAAAAAGAGAGAAACACCGGTTGTGGCT 1500  
Db 1416 CCCGCTGAGAGTGTGATTAATCTCTTCAAGAGAAAAAGAGAGAAACACCGGTTGTGGCT 1475  
QY 1501 GCACACTGCAAGAAAGATACAGCTGAAAAAGAGCGGCTCTTAAACAGTTTACACTAT 1560  
Db 1476 GCACACTGCAAGAAAGATACAGCTGAAAAAGAGCGGCTCTTAAACAGTTTACACTAT 1535  
QY 1561 CAACCGTGTATCATCAAGGAGCGCTGTGACAGTTGTCGCTTGTGTATAGACAA 1620  
Db 1536 CAACCGTGTATCATCAAGGAGCGCTGTGACAGTTGTCGCTTGTGTATAGACAA 1595  
QY 1621 AATTTTGTGAAAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAACCGCTTTCCGGAGTGC 1680  
Db 1596 AATTTTGTGAAAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAACCGCTTTCCGGAGTGC 1655  
QY 1681 CGCTGCAAGCAAGTGTGCAACCAAGAGTCCGCTGTACTGTGCTGTCCGAGAGTGT 1740  
Db 1656 CGCTGCAAGCAAGTGTGCAACCAAGAGTCCGCTGTACTGTGCTGTCCGAGAGTGT 1715  
QY 1741 GACCCTGACTCTGTCTTACTTGTGTGAGCGCTGACCAATGGGACATGAAAAATGTGTC 1800  
Db 1716 GACCCTGACTCTGTCTTACTTGTGTGAGCGCTGACCAATGGGACATGAAAAATGTGTC 1775  
QY 1801 TGCAAGAACTGCAAGTATCAAGGGGGCTCCAAAAAGCATCTATGCTGTGCAACATCTGAC 1860  
Db 1776 TGCAAGAACTGCAAGTATCAAGGGGGCTCCAAAAAGCATCTATGCTGTGCAACATCTGAC 1835  
QY 1861 GTGGAGGCTGGGGATTTTATCAAGATCCTGTGCAAGAAAAATGAATTCATCTCAGAA 1920  
Db 1836 GTGGAGGCTGGGGATTTTATCAAGATCCTGTGCAAGAAAAATGAATTCATCTCAGAA 1895  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGGAGAAAGTGTATGATTA 1980  
Db 1896 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGGAGAAAGTGTATGATTA 1955  
QY 1981 TACATGTGCAAGTATCTGTCAACTTGAACTGATTTGTGTGATGCAACCCGCAAG 2040  
Db 1956 TACATGTGCAAGTATCTGTCAACTTGAACTGATTTGTGTGATGCAACCCGCAAG 2015  
QY 2041 GGTAACTAAATTCGTTTGGCAATGATTCGGTAAATCCAAACTGTATGCAAAATTTG 2100  
Db 2016 GGTAACTAAATTCGTTTGGCAATGATTCGGTAAATCCAAACTGTATGCAAAATTTG 2075  
QY 2101 ATGTGTAACGAGTATCAAGATAGATTAATTTTGCAGAGAGCCATCCAGCTGGCA 2160  
Db 2076 ATGTGTAACGAGTATCAAGATAGATTAATTTTGCAGAGAGCCATCCAGCTGGCA 2135  
QY 2161 GAGCTGTTTTTGTATTACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGCATGAA 2220  
Db 2136 GAGCTGTTTTTGTATTACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGCATGAA 2195  
QY 2221 AGAGAAATGGAATCCCTTGA 2241  
Db 2196 AGAGAAATGGAATCCCTTGA 2216

RESULT 13  
AX747667 2476 bp mRNA linear PAT 20-JUN-2003  
DEFINITION Sequence 1192 from Patent EP1308459.  
ACCESSION AX747667  
VERSION AX747667.1 GI:32132055  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE  
1  
Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and  
Masuno, Y.  
Full-length cDNA sequences  
Patent: EP 1308459-A 1192 07-MAY-2003;  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)  
FEATURES  
Location/Qualifiers  
source  
1..2476  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 89.0%; Score 1993.8; DB 2; Length 2476;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 2; Indels 117; Gaps 1;  
QY 1 ATGGGCGCAGCTGGGAGGAATCTGAGAGGAGCACGTTGTTGGCGGAGCGGTAAAA 60  
Db 93 ATGGGCGCAGCTGGGAGGAATCTGAGAGGAGCACGTTGTTGGCGGAGCGGTAAAA 152  
QY 61 TCAGAGTACATCGACTGAGACAGCTCAAGAGTTTCAAGACGCTGATGAAAGAGT 120  
Db 153 TCAGAGTACATCGACTGAGACAGCTCAAGAGTTTCAAGACGCTGATGAAAGAGT 212  
QY 121 ATGTTTATGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAACCAAGATGG 180  
Db 213 ATGTTTATGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAACCAAGATGG 272  
QY 181 AAACAGGAGAGATACAGCTGTGACATCTGTGACTTCTGTGAGCTCATTTGGCGGAGCT 240  
Db 273 AAACAGGAGAGATACAGCTGTGACATCTGTGACTTCTGTGAGCTCATTTGGCGGAGCT 332  
QY 241 AGGAGTGTTCGATGACCAAGTGAATTTTCAACACAAATCATCCATTAAAGACT 300  
Db 333 AGGAGTGTTCGATGACCAAGTGAATTTTCAACACAAATCATCCATTAAAGACT 337  
QY 301 CTGAATGCAAGTTGCTTCAATGATCAATTAATGATTTCTGCTCCCTTACAGCAATTTT 360  
Db 338 ----- 337  
QY 361 ATGTGGAAGATGAACGTTTTCATTAACATTCCTTATATGAGAGATGAAGTTTATGAT 420  
Db 338 --GGTGAAGATGAACGTTTTCATTAACATTCCTTATATGAGAGATGAAGTTTATGAT 395  
QY 421 CAGAGTGTACTTTCATTGAAAGAACTAATAAAAAATTAATGATGGAAGATACAGGGAGT 480  
Db 396 CAGAGTGTACTTTCATTGAAAGAACTAATAAAAAATTAATGATGGAAGATACAGGGAGT 455  
QY 481 AGAGAAATGTGGGTTTAAATGAAGAAATTTTGTGAGTGTGTGATGCCCTTGTGCA 540  
Db 456 AGAGAAATGTGGGTTTAAATGAAGAAATTTTGTGAGTGTGTGATGCCCTTGTGCA 515  
QY 541 TATAATGATGATGACATGATGATGATGAGACAGATCCTGAAGAAAGAGAAAGAGAG 600  
Db 516 TATAATGATGATGACATGATGATGATGAGACAGATCCTGAAGAAAGAGAAAGAGAG 575

QY 601 AAAGATCTGGAGATCAACGAGATGATAAGAAAGCCGCCACTTCGAAATTTCTCT 660  
 Db 576 AAAGATCTGGAGATCAACGAGATGATAAGAAAGCCGCCACTTCGAAATTTCTCT 635  
 QY 661 GATATAATTTTGAAGCCATTTCTCAATGTTTCCAGATAAGGCAAGCAGAGAACTA 720  
 Db 636 GATATAATTTTGAAGCCATTTCTCAATGTTTCCAGATAAGGCAAGCAGAGAACTA 695  
 QY 721 AAGAAAAATATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGATGT 780  
 Db 636 AAGAAAAATATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGATGT 755  
 QY 781 ACCCCCAATGATGAGACCAAAATGCTAAATCTGTTGAGAGAGAGAAAGCTTACATCC 840  
 Db 756 ACCCCCAATGATGAGACCAAAATGCTAAATCTGTTGAGAGAGAGAAAGCTTACATCC 815  
 QY 841 TTTCAATGCTTTCTGTAGGCGATGTTTAAATATGACTGCTCTCAATCTTTTCAT 900  
 Db 816 TTTCAATGCTTTCTGTAGGCGATGTTTAAATATGACTGCTCTCAATCTTTTCAT 875  
 QY 901 GCAACACCCCAACATTAATAGCGGAGAAACACAGAAACAGCTCTAGACAAACCTTGT 960  
 Db 876 GCAACACCCCAACATTAATAGCGGAGAAACACAGAAACAGCTCTAGACAAACCTTGT 935  
 QY 961 GGAACCAAGTGTACCAAGATTTTGAAGGAGCAAAAGAGTTTGTCTGTCTTCAACGCT 1020  
 Db 936 GGAACCAAGTGTACCAAGATTTTGAAGGAGCAAAAGAGTTTGTCTGTCTTCAACGCT 995  
 QY 1021 GAGCGGATTAAGACCCCAACAAACGTCAGAGAGCGCGAGAAAGAGAGCGCTTCCAT 1080  
 Db 996 GAGCGGATTAAGACCCCAACAAACGTCAGAGAGCGCGAGAAAGAGAGCGCTTCCAT 1055  
 QY 1081 AAGAGTACAGGCGCCAGACACCCCAACATTAATGTGTGATCAAGAGATACAGAGT 1140  
 Db 1056 AAGAGTACAGGCGCCAGACACCCCAACATTAATGTGTGATCAAGAGATACAGAGT 1115  
 QY 1141 GATAGGAG 1200  
 Db 1116 GATAGGAG 1175  
 QY 1201 AAAGATGAACCTTCGAGCTCTCTGAGAGCAATTTCTGCTGTCGCAACCAATTAAGATG 1260  
 Db 1176 AAAGATGAACCTTCGAGCTCTCTGAGAGCAATTTCTGCTGTCGCAACCAATTAAGATG 1235  
 QY 1261 AAGCCAAATATTTGAACCTTCGAGAGATGTGAGAGTGTCTGAAGCCTCAATGTTT 1320  
 Db 1236 AAGCCAAATATTTGAACCTTCGAGAGATGTGAGAGTGTCTGAAGCCTCAATGTTT 1295  
 QY 1321 AGAGTCTCATTTGGAAGCTTCTGAGATGTGAGAGTGTCTGAAGCCTCAATGTTT 1380  
 Db 1296 AGAGTCTCATTTGGAAGCTTCTGAGATGTGAGAGTGTCTGAAGCCTCAATGTTT 1355  
 QY 1381 AAAACATGTAGACAGGTGTATGAGTTTGAATCAAGAAATCTAGATCATAGCTCAGCT 1440  
 Db 1356 AAAACATGTAGACAGGTGTATGAGTTTGAATCAAGAAATCTAGATCATAGCTCAGCT 1415  
 QY 1441 CCGGCTGAGATGTGATGATCTCTCCAGAGAAAGAAAGAGAGAAACACCGGTTGTGGCT 1500  
 Db 1416 CCGGCTGAGATGTGATGATCTCTCCAGAGAAAGAAAGAGAGAAACACCGGTTGTGGCT 1475  
 QY 1501 GCACTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
 Db 1476 GCACTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1535  
 QY 1561 CAACCTGTGATCATCAACGAGAGCTGTGACAGTGTGCTGTGTGTATGATGATGAT 1620  
 Db 1536 CAACCTGTGATCATCAACGAGAGCTGTGACAGTGTGCTGTGTGTATGATGATGAT 1595  
 QY 1621 AATTTTGTGAAGATTTTGTCAATGTATGTGATGATGATGATGATGATGATGATGAT 1680  
 Db 1596 AATTTTGTGAAGATTTTGTCAATGTATGTGATGATGATGATGATGATGATGATGAT 1655

QY 1681 CGCTGCAAGAGACAGTGTGACACCAAGCAGTGCCTGCTCACTGCTGTCCAGAGTGT 1740  
 Db 1656 CGCTGCAAGAGACAGTGTGACACCAAGCAGTGCCTGCTCACTGCTGTCCAGAGTGT 1715  
 QY 1741 GACCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATGGGACATGATGATGATGAT 1800  
 Db 1716 GACCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATGGGACATGATGATGATGAT 1775  
 QY 1801 TGCAGAGACCTGAGATTTCAAGGCGGCTCCAAAAGAGATCTTATGCTGTGACATCTGAC 1860  
 Db 1776 TGCAGAGACCTGAGATTTCAAGGCGGCTCCAAAAGAGATCTTATGCTGTGACATCTGAC 1835  
 QY 1861 GTGGCAGGCTGGGGATTTTATCAAGATCTGTGAGAGAAATGATGATGATGATGATGAT 1920  
 Db 1836 GTGGCAGGCTGGGGATTTTATCAAGATCTGTGAGAGAAATGATGATGATGATGATGAT 1895  
 QY 1921 TACTGTGAGAGATTTTCTCAAGATGAGCTGACAGAGAGAGAGAGAGAGATGATGATGAT 1980  
 Db 1896 TACTGTGAGAGATTTTCTCAAGATGAGCTGACAGAGAGAGAGAGAGAGATGATGATGAT 1955  
 QY 1981 TACATGTGAGCTTTCTGTCTTCACTTGAACATGATTTTGTGTGTGATGATGATGATGAT 2040  
 Db 1956 TACATGTGAGCTTTCTGTCTTCACTTGAACATGATTTTGTGTGTGATGATGATGATGAT 2015  
 QY 2041 GGTAAACAAATTTGTTTGAATCAATCATTCGTTAATCCAACTGCTATGCAAAAGTTATG 2100  
 Db 2016 GGTAAACAAATTTGTTTGAATCAATCATTCGTTAATCCAACTGCTATGCAAAAGTTATG 2075  
 QY 2101 ATGTGTTAACGCTGATCAAGATGATGATTTTGTGCAAGAGAGCCATGACAGTGGCGAA 2160  
 Db 2076 ATGTGTTAACGCTGATCAAGATGATGATTTTGTGCAAGAGAGCCATGACAGTGGCGAA 2135  
 QY 2161 GAGCTGTTTTTGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2220  
 Db 2136 GAGCTGTTTTTGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2195  
 QY 2221 AGAGAAATGAGAAATCCCTTGA 2241  
 Db 2196 AGAGAAATGAGAAATCCCTTGA 2216

RESULT 14  
 AK092676  
 LOCUS 2476 bp mRNA linear PRI 20-JAN-2006  
 DEFINITION Homo sapiens cDNA FLJ35357 fis, clone FUAEN200312, highly similar  
 to ENHANCER OF ZEST2 HOMOLOG 2;  
 ACCESSION AK092676  
 VERSION AK092676.1 GI:21751325  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

# REFERENCE AUTHORS

1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
 Sekine, M., Ohtsuka, M., Nishitani, T., Shibahara, T., Tanaka, T.,  
 Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
 Nagahara, K., Murakami, K., Yasuda, T., Iwiyama, T., Magatsuma, M.,  
 Shiraori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
 Kikkawa, B., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
 Tanikawa, M., Yamazaki, M., Niimoriya, K., Ishibashi, T., Yamashita, H.,  
 Hirakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
 Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
 Yosida, M., Hoshino, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,  
 Hara, H., Tanase, T.O., Nomura, Y., Togiya, S., Komai, F., Hara, A.,  
 Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A.,  
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
 Shiohata, N., Sano, S., Moriyama, S., Moriyama, H., Sato, N., Takami, S.,  
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujisawa, T., Oho, T., Yamada, K., Fujii, Y.,  
Ozaki, K., Hirao, M., Ohnori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senda, T., Matsuura, K.,  
Nakajima, Y., Mizuno, T., Morinaga, M., Saeaki, M., Togaishi, T.,  
Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamauchi, R.,  
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegai, T. and Sugano, S.  
Complete sequencing and characterization of 21,243 full-length  
human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

2  
Ishibashi, T., Kanehori, K., Yoshida, M., Watanabe, S., Ishida, S.,  
Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,  
Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,  
Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,  
Makemitsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Sato, K.,  
Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,  
Sethine, M., Kikuchi, H., Kanda, K., Nagatsuma, M., Takahashi, F.,  
Oshima, M., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagaharti, K., Masuh, Y., Nagai, K. and Isegai, T.  
NEBO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2476)  
Isegai, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isegai, Helix Research Institute,  
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
E-mail: f1j-cdn@helix.com, Tel: 81-438-52-3975, Fax: 81-438-52-3986  
NEBO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan, cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
Location/Qualifiers  
1..2476  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="PUBEN2000312"  
/cell\_type="pulmonary artery endothelial cells (HPAEC)"  
/clone\_id="PUBEN2"  
/note="cloning vector: pME18SFL3  
primary culture, pulmonary artery endothelial cells"

ORIGIN

Query Match 89.0%; Score 1993.8; DB 5; Length 2476;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 2; Indels 117; Gaps 1;

QY 1 ATGGGCGAGCTGGGAGAGAAATCTGAGAGGGAGCAAGTTTGGCGGAAGGCTGTAATA 60  
DB 93 ATGGGCGAGCTGGGAGAGAAATCTGAGAGGGAGCAAGTTTGGCGGAAGGCTGTAATA 152  
QY 61 TCAGAGTACATGCGACGTGAGACAGCTCAAGAGTTTCAAGCGAGCTGATGAAGTAAGAGT 120  
DB 153 TCAGAGTACATGCGACGTGAGACAGCTCAAGAGTTTCAAGCGAGCTGATGAAGTAAGT 212  
QY 121 ATGTTTAACTTCAATGCTGAGAAATTTTGGAAAGAGGAATTTTAAACCAAGAAATGG 180  
DB 213 ATGTTTAACTTCAATGCTGAGAAATTTTGGAAAGAGGAATTTTAAACCAAGAAATGG 272  
QY 181 AAACGAGGAAGGATCAGCGCTGTGACATCTGAGCTTCTGAGAGCTCATTTGGCGGAGACT 240  
DB 273 AAACGAGGAAGGATCAGCGCTGTGACATCTGAGCTTCTGAGAGCTCATTTGGCGGAGACT 332

QY 241 AGGAGTCTTCGGTGACCACTGACTTGATTTTCCAAACAGATCATCCATTAAAGACT 300  
DB 333 AGGGA----- 337  
QY 301 CTGAATGCACTTGCTTCAGTACCAATATATGATCTTGCTCCCTTACAGACAAATTTT 360  
DB 338 ----- 337  
QY 361 ATGTGTGAAGATGAAACCTGTTTATCATTAATCTTATATGAGATGAAAGTTTATGAT 420  
DB 338 --GGTGAAGATGAAATGTTTTATCATTAATCTTATATGAGATGAAAGTTTATGAT 395  
QY 421 CAGATGTACTTTCATTGGAAGAACTATATATATATATATATATATATATATATATAT 480  
DB 396 CAGATGTACTTTCATTGGAAGAACTATATATATATATATATATATATATATATATAT 455  
QY 481 AGAGATGTGGGTTTAT 540  
DB 456 AGAGATGTGGGTTTAT 515  
QY 541 TAT 600  
DB 516 TAT 575  
QY 601 AAAGATCTGAGATGATCCAGATGATATATATATATATATATATATATATATATATAT 660  
DB 576 AAAGATCTGAGATGATCCAGATGATATATATATATATATATATATATATATATATAT 635  
QY 661 GATTAATATTTTGAAGCATTTCTCTCAATGTTTCCATATATATATATATATATATATAT 720  
DB 636 GATTAATATTTTGAAGCATTTCTCTCAATGTTTCCATATATATATATATATATATATAT 695  
QY 721 AAGGAAAT 780  
DB 696 AAGGAAAT 755  
QY 781 ACCCCCAAT 840  
DB 756 ACCCCCAAT 815  
QY 841 TTTTATATGCTTTCTGTAAGCGATGTTTAAATATATATATATATATATATATATATAT 900  
DB 816 TTTTATATGCTTTCTGTAAGCGATGTTTAAATATATATATATATATATATATATATAT 875  
QY 901 GCAACACCCCAAT 960  
DB 876 GCAACACCCCAAT 935  
QY 961 GGATCCAGATGTTTACAGATATTTGGAGAGCAAGAGTTTGGTGTCTTCAACGGCT 1020  
DB 936 GGATCCAGATGTTTACAGATATTTGGAGAGCAAGAGTTTGGTGTCTTCAACGGCT 995  
QY 1021 GAGCGATATTAAGACCCCAACCAAACTGTCAGAGAGCGCAGAAAGAGAGCGCTTCCCAAT 1080  
DB 996 GAGCGATATTAAGACCCCAACCAAACTGTCAGAGAGCGCAGAAAGAGAGCGCTTCCCAAT 1055  
QY 1081 AACATGACAGAGCCCAAGACCCCAACCAATATATATATATATATATATATATATATAT 1140  
DB 1056 AACATGACAGAGCCCAAGACCCCAACCAATATATATATATATATATATATATATATAT 1115  
QY 1141 GATGAGGAGAGGAGATGAAACGGGGGAGAGAACTATATATATATATATATATATATAT 1200  
DB 1116 GATGAGGAGAGGAGATGAAACGGGGGAGAGAACTATATATATATATATATATATATAT 1175  
QY 1201 AAAGATGAACTTGAAGCTCTCTGAAGCAAAATCTCGGTCTCAACACCAATATATATAT 1260  
DB 1176 AAAGATGAACTTGAAGCTCTCTGAAGCAAAATCTCGGTCTCAACACCAATATATATAT 1235  
QY 1261 AAGCAAT 1320  
DB 1236 AAGCAAT 1295  
QY 1321 AGAGTCTCATTTGACATTAATGACAAATTTCTGTGCAATTTGCTAGATTAAATTGGAGCC 1380











CC The DNA was isolated by screening a human B cell cDNA library with mixed  
CC Drosophila DNA probes based on the conserved SET domains in E(2) and  
CC Su(Var)3-9. The DNA, and its products, are useful in therapy (esp. gene  
CC therapy) and diagnosis of human diseases that involve deregulated  
CC chromatin-regulator genes having a SET domain, esp. cancer. (Updated on  
CC 25-MAR-2003 to correct PR field.)  
XX

Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

Query Match 100.0%; Score 2241; DB 2; Length 2600;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGCCGACGCTGGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGACGCTGTAAA 60
DB ATGGCCGACGCTGGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGAGCGGTAAA 149
QY 61 TCAGAGTACATGCGACCTGAGACAGCTCAAGAGTTTCAGACGAGCTGATGAAAGAGT 120
DB TCAGAGTACATGCGACCTGAGACAGCTCAAGAGTTTCAGACGAGCTGATGAAAGAGT 209
QY 121 ATGTTTGGTCCATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGAAATG 180
DB ATGTTTGGTCCATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGAAATG 269
QY 210 ATGTTTGGTCCATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGAAATG 269
DB ATGTTTGGTCCATCGTCAGAAAATTTGGAAAAGACGAAATCTTAAACCAAGAAATG 269
QY 181 AAACGAGAGAGATACAGCTGTGCACTCTGACTTCTGTAGCTCATTCGCGGAGCT 240
DB AAACGAGAGAGATACAGCTGTGCACTCTGACTTCTGTAGCTCATTCGCGGAGCT 240
QY 270 AAACGAGAGAGATACAGCTGTGCACTCTGACTTCTGTAGCTCATTCGCGGAGCT 329
DB AAACGAGAGAGATACAGCTGTGCACTCTGACTTCTGTAGCTCATTCGCGGAGCT 329
QY 241 AGGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 300
DB AGGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 389
QY 330 AGGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 389
DB AGGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 389
QY 301 CTGATGAGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 360
DB CTGATGAGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 449
QY 390 CTGATGAGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 449
DB CTGATGAGAGTGTTCGGTGAACGAGTCTGGATTTTCCACACAGTCAATCCCTTAAAGACT 449
QY 361 ATGTTGAGAGATGAACTGTTTTCATTAACATTCCTTATATGAGATGAAAGTTTATAT 420
DB ATGTTGAGAGATGAACTGTTTTCATTAACATTCCTTATATGAGATGAAAGTTTATAT 509
QY 450 ATGTTGAGAGATGAACTGTTTTCATTAACATTCCTTATATGAGATGAAAGTTTATAT 509
DB ATGTTGAGAGATGAACTGTTTTCATTAACATTCCTTATATGAGATGAAAGTTTATAT 509
QY 421 CAGAGTGTTCCTTATGAAAGAACTAATAAAATTTATGATGGAAAGTACAGGGAGT 480
DB CAGAGTGTTCCTTATGAAAGAACTAATAAAATTTATGATGGAAAGTACAGGGAGT 569
QY 510 CAGAGTGTTCCTTATGAAAGAACTAATAAAATTTATGATGGAAAGTACAGGGAGT 569
DB CAGAGTGTTCCTTATGAAAGAACTAATAAAATTTATGATGGAAAGTACAGGGAGT 569
QY 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGATTTGGTGAATGCCCTGTGCA 540
DB AGAGATGTGGGTTTATTAATGATGAATTTTGTGGATTTGGTGAATGCCCTGTGCA 629
QY 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGATTTGGTGAATGCCCTGTGCA 629
DB AGAGATGTGGGTTTATTAATGATGAATTTTGTGGATTTGGTGAATGCCCTGTGCA 629
QY 541 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
QY 630 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
DB TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
QY 601 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660
DB AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749
QY 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749
DB AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749
QY 661 GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 720
DB GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 809
QY 750 GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 809
DB GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACAGCAAGAACTA 809
QY 721 AAGGAAAAATATAAGAACTCACCGAAGCAGCTCCGAGGCGCACTTCTCTGAAATGT 780
DB AAGGAAAAATATAAGAACTCACCGAAGCAGCTCCGAGGCGCACTTCTCTGAAATGT 869
QY 810 AAGGAAAAATATAAGAACTCACCGAAGCAGCTCCGAGGCGCACTTCTCTGAAATGT 869
DB AAGGAAAAATATAAGAACTCACCGAAGCAGCTCCGAGGCGCACTTCTCTGAAATGT 869
QY 781 ACCCCCAATATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840
DB ACCCCCAATATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 929
QY 870 ACCCCCAATATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 929
DB ACCCCCAATATGATGACCAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 929
QY 841 TTTGATAGCTTTTCTGATAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 900
DB TTTGATAGCTTTTCTGATAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 989
QY 930 TTTGATAGCTTTTCTGATAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 989
DB TTTGATAGCTTTTCTGATAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 989
```

```
QY 901 GCACACCCCAACCTTATTAAGCCGAGAGACACGAAACAGCTCTACACAAACCTTGT 960
DB GCACACCCCAACCTTATTAAGCCGAGAGACACGAAACAGCTCTACACAAACCTTGT 1049
QY 990 GCACACCCCAACCTTATTAAGCCGAGAGACACGAAACAGCTCTACACAAACCTTGT 1049
DB GCACACCCCAACCTTATTAAGCCGAGAGACACGAAACAGCTCTACACAAACCTTGT 1049
QY 961 GAGACACAGGTTTACAGCACTTTGAGGGGAGCAAAAGGATTTGTGCTGCTCACCGCT 1020
DB GAGACACAGGTTTACAGCACTTTGAGGGGAGCAAAAGGATTTGTGCTGCTCACCGCT 1109
QY 1050 GAGACACAGGTTTACAGCACTTTGAGGGGAGCAAAAGGATTTGTGCTGCTCACCGCT 1109
DB GAGACACAGGTTTACAGCACTTTGAGGGGAGCAAAAGGATTTGTGCTGCTCACCGCT 1109
QY 1021 GAGCGGATTAAGACCCCAACCAACCTGTCAGAGGCGCGCAAGAGAGAGCGCTTCCCAAT 1080
DB GAGCGGATTAAGACCCCAACCAACCTGTCAGAGGCGCGCAAGAGAGAGCGCTTCCCAAT 1169
QY 1110 GAGCGGATTAAGACCCCAACCAACCTGTCAGAGGCGCGCAAGAGAGAGCGCTTCCCAAT 1169
DB GAGCGGATTAAGACCCCAACCAACCTGTCAGAGGCGCGCAAGAGAGAGCGCTTCCCAAT 1169
QY 1081 AACATGACAGGCGCGCACACCCCAACCAATTAATGTGTGAAATCAAAAGATACAGAGT 1140
DB AACATGACAGGCGCGCACACCCCAACCAATTAATGTGTGAAATCAAAAGATACAGAGT 1229
QY 1170 AACATGACAGGCGCGCACACCCCAACCAATTAATGTGTGAAATCAAAAGATACAGAGT 1229
DB AACATGACAGGCGCGCACACCCCAACCAATTAATGTGTGAAATCAAAAGATACAGAGT 1229
QY 1141 GATGAGGAGACAGGAGCTGAAAAGGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1200
DB GATGAGGAGACAGGAGCTGAAAAGGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1289
QY 1230 GATGAGGAGACAGGAGCTGAAAAGGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1289
DB GATGAGGAGACAGGAGCTGAAAAGGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1289
QY 1201 AAAGATGAACTTCGAGCTCCTGTAAGCAAAATTCGTGTCAAAACCAATTAAGATG 1260
DB AAAGATGAACTTCGAGCTCCTGTAAGCAAAATTCGTGTCAAAACCAATTAAGATG 1349
QY 1290 AAAGATGAACTTCGAGCTCCTGTAAGCAAAATTCGTGTCAAAACCAATTAAGATG 1349
DB AAAGATGAACTTCGAGCTCCTGTAAGCAAAATTCGTGTCAAAACCAATTAAGATG 1349
QY 1261 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGATGATGATGATGATGATGAT 1320
DB AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGATGATGATGATGATGATGAT 1409
QY 1350 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGATGATGATGATGATGATGAT 1409
DB AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGATGATGATGATGATGATGAT 1409
QY 1321 AAGGCTCTCAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
DB AAGGCTCTCAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1469
QY 1410 AAGGCTCTCAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1469
DB AAGGCTCTCAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1469
QY 1381 AAAACATGATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB AAAACATGATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529
QY 1470 AAAACATGATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529
DB AAAACATGATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529
QY 1441 CCGGCTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB CCGGCTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589
QY 1530 CCGGCTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589
DB CCGGCTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589
QY 1501 GCACACTGCAAGAAAGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1560
DB GCACACTGCAAGAAAGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1649
QY 1590 GCACACTGCAAGAAAGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1649
DB GCACACTGCAAGAAAGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1649
QY 1561 CAACCTGTGATCATCCACGAGCCTGTGACAGTTGATGATGATGATGATGATGATGATGAT 1620
DB CAACCTGTGATCATCCACGAGCCTGTGACAGTTGATGATGATGATGATGATGATGATGAT 1709
QY 1650 CAACCTGTGATCATCCACGAGCCTGTGACAGTTGATGATGATGATGATGATGATGATGAT 1709
DB CAACCTGTGATCATCCACGAGCCTGTGACAGTTGATGATGATGATGATGATGATGATGAT 1709
QY 1621 AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
DB AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
QY 1710 AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
DB AATTTTGTGAAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
QY 1681 CGCTGCAAGAGCAGTGAACACCAAGCAAGTGCCTGTGATGATGATGATGATGATGATGAT 1740
DB CGCTGCAAGAGCAGTGAACACCAAGCAAGTGCCTGTGATGATGATGATGATGATGATGAT 1829
QY 1770 CGCTGCAAGAGCAGTGAACACCAAGCAAGTGCCTGTGATGATGATGATGATGATGATGAT 1829
DB CGCTGCAAGAGCAGTGAACACCAAGCAAGTGCCTGTGATGATGATGATGATGATGATGAT 1829
QY 1741 GACCTGACCTCTGTCTTATCTTGTGAGCGGCTGACCAATGGGACAGTAAATGTGTCC 1800
DB GACCTGACCTCTGTCTTATCTTGTGAGCGGCTGACCAATGGGACAGTAAATGTGTCC 1889
QY 1830 GACCTGACCTCTGTCTTATCTTGTGAGCGGCTGACCAATGGGACAGTAAATGTGTCC 1889
DB GACCTGACCTCTGTCTTATCTTGTGAGCGGCTGACCAATGGGACAGTAAATGTGTCC 1889
QY 1801 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCAATATGCTGCGACATCTGAC 1860
DB TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCAATATGCTGCGACATCTGAC 1949
QY 1890 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCAATATGCTGCGACATCTGAC 1949
DB TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCAATATGCTGCGACATCTGAC 1949
QY 1861 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGATATCTCAAGAA 1920
DB GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGATATCTCAAGAA 2009
QY 1950 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGATATCTCAAGAA 2009
DB GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGATATCTCAAGAA 2009
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 1980
DB TACTGTGAGAGATTAATTTCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 2069
QY 2010 TACTGTGAGAGATTAATTTCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 2069
DB TACTGTGAGAGATTAATTTCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 2069
QY 1981 TACATGTGACGCTTTCTGTTCACCTGAAACATGATTTTGTGTGATGACACCGGCAAG 2040
DB TACATGTGACGCTTTCTGTTCACCTGAAACATGATTTTGTGTGATGACACCGGCAAG 2040
```

DB 2070 TACAGTGTGAGCTTTCTGTCTACTGTAACATGATTTGTGTGATGATCAACCGCAG 2129  
QY 2041 GGTAAACAAATTCGTTTSCAAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2100  
DB 2130 GGTAAACAAATTCGTTTSCAAATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2189  
QY 2101 ATGTTAAGCGTATGATCAAGATAGTATTTTGGCAAGAGCCATCCAGCTGCGCA 2160  
DB 2190 ATGTTAAGCGTATGATCAAGATAGTATTTTGGCAAGAGCCATCCAGCTGCGCA 2249  
QY 2161 GAGCTGTTTGTGATTAACATATACAGCCAGCTGATCCCTGAAGTATGCGCATCGAA 2220  
DB 2250 GAGCTGTTTGTGATTAACATATACAGCCAGCTGATCCCTGAAGTATGCGCATCGAA 2309  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
DB 2310 AGAGAAATGAAATCCCTTGA 2330

RESULT 2  
ABL65847  
ID ABL65847 standard; DNA; 2600 BP.

AC ABL65847;  
XX  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Lung cancer related gene sequence SEQ ID NO:4184.  
XX  
XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;  
KW gene; ds.

OS Homo sapiens.

PN WO200194629-A2.

XX 13-DEC-2001.

PF 30-MAY-2001; 2001WO-US010838.

XX  
PR 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 28-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.

PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.

XX (AVAL-) AVALON PHARM.

XX PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;

XX WPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 4184; 44bp; English.

XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (1) of a signature gene set, where (1)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (1) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
CC tumour

XX Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

Query Match 100.0%; Score 2241; DB 6; Length 2600;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCCAGACTGGAGAGAAATCTGAGAGGAGCAGTTTGTGGCGGAGCGTATAAA 60  
DB 90 ATGGGCCAGACTGGAGAGAAATCTGAGAGGAGCAGTTTGTGGCGGAGCGTATAAA 149  
QY 61 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAGACGATGAGTAAGAGT 120  
DB 150 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAGACGATGAGTAAGAGT 209  
QY 121 ATGTTAGTTCATATCGTCAGAAAATTTGGAAAAGACGAATCTTAACCAAGATG 180  
DB 210 ATGTTAGTTCATATCGTCAGAAAATTTGGAAAAGACGAATCTTAACCAAGATG 269  
QY 181 AAACAGCAGAGATPACAGCTGTGACATCTGTGAGCTCATTTGGCGGAGCT 240  
DB 270 AAACAGCAGAGATPACAGCTGTGACATCTGTGAGCTCATTTGGCGGAGCT 329  
QY 241 AGGAGCTTGGGAGACAGTGAATTTTCCACACAAAGTATCCATTAAGACT 300  
DB 330 AGGAGCTTGGGAGACAGTGAATTTTCCACACAAAGTATCCATTAAGACT 389  
QY 301 CTGAATGAGTTGCTTCAATGATCAATATGATTTCTGCTCCCTACAGCAATTTT 360  
DB 390 CTGAATGAGTTGCTTCAATGATCAATATGATTTCTGCTCCCTACAGCAATTTT 449

OY	361	ATGTGGAAATGAAACGTGTTTACATPACATTCCTTAATATGGAGATGAAGTTTATGAT	420
Db	450	ATGTGGAAATGAAACGTGTTTACATPACATTCCTTAATATGGAGATGAAGTTTATGAT	509
OY	421	CAGATGGTACTTTCATTGAAGAACTATATAAAAAATTATGATGGAAAGTACACGGGAT	480
Db	510	CAGATGGTACTTTCATTGAAGAACTATATAAAAAATTATGATGGAAAGTACACGGGAT	569
OY	481	AGAGAATGTGGCTTTATTAATGATGAATTTTGTGGAGTTGGTGAATGCCCCTTGGTCAA	540
Db	570	AGAGAATGTGGCTTTATTAATGATGAATTTTGTGGAGTTGGTGAATGCCCCTTGGTCAA	629
OY	541	TATATATGATGATGACGATGATGATGATGGAACGATCCTGAAAGAAAGAAAGAACAG	600
Db	630	TATATATGATGATGACGATGATGATGATGGAACGATCCTGAAAGAAAGAAAGAACAG	689
OY	601	AAAGATCTGGAGATGACCCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCTCT	660
Db	690	AAAGATCTGGAGATGACCCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCTCT	749
OY	661	GATTAATATTTTGAAGCCATTTTCTCATGTTTTCAGATTAAGGGCACAGCAAGAAACTA	720
Db	750	GATTAATATTTTGAAGCCATTTTCTCATGTTTTCAGATTAAGGGCACAGCAAGAAACTA	809
OY	721	AAGGAAAAATTAATAAGAACTACCGAAACGAGCGTCCGACGCACTTCTCTGTAATGT	780
Db	810	AAGGAAAAATTAATAAGAACTACCGAAACGAGCGTCCGACGCACTTCTCTGTAATGT	869
OY	781	ACCCCAACATATAGATGACCAATGCTAAATCTGTTCAAGAGACCAAGCTTACACTCC	840
Db	870	ACCCCAACATATAGATGACCAATGCTAAATCTGTTCAAGAGACCAAGCTTACACTCC	929
OY	841	TTTCATACGCTTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTACATCTTTTCAT	900
Db	930	TTTCATACGCTTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTACATCTTTTCAT	989
OY	901	GCAACACCAACACTTATATAGCGGAGAAACACAGAAACGCTTACAGCAACAACTTGT	960
Db	990	GCAACACCAACACTTATATAGCGGAGAAACACAGAAACGCTTACAGCAACAACTTGT	1049
OY	961	GGACCAACAGTGTATACAGCAATTTGGAGGGACAAAGATTTGCTGCTCTCACCGCT	1020
Db	1050	GGACCAACAGTGTATACAGCAATTTGGAGGGAGAAAGATTTGCTGCTCTCACCGCT	1109
OY	1021	GAGCGGATTAAGACCCCAACCAAACTCCAGAGGCGCGAAGAGACGCGTTCCCAAT	1080
Db	1110	GAGCGGATTAAGACCCCAACCAAACTCCAGAGGCGCGAAGAGACGCGTTCCCAAT	1169
OY	1081	AAACGTACAGGCCCCAGACACCCCACTTAATGTGCTGGAAATCAAGATATACAGACGT	1140
Db	1170	AAACGTACAGGCCCCAGACACCCCACTTAATGTGCTGGAAATCAAGATATACAGACGT	1229
OY	1141	GATAGGGAACAGGGACTGAAGACGGGGGGAAGAACTGATTAAGAAAGAAAGAGAG	1200
Db	1230	GATAGGGAACAGGGACTGAAGACGGGGGGAAGAACTGATTAAGAAAGAAAGAGAG	1289
OY	1201	AAAGATGAACCTTGCAGCTCCTCTGAGCAAAATTTCGCTGCAAAACCAATTAAGATG	1260
Db	1290	AAAGATGAACCTTGCAGCTCCTCTGAGCAAAATTTCGCTGCAAAACCAATTAAGATG	1349
OY	1261	AAAGCAATATTAAGACCTCCGAGAAATGTGGATGTGGATGTGGAGCGTCAATGTTT	1330
Db	1350	AAAGCAATATTAAGACCTCCGAGAAATGTGGATGTGGATGTGGAGCGTCAATGTTT	1409
OY	1321	AGAGTCTCATTTGGCACTTATCTATGACAAATTTCTGTGCATTTGCTAGGTTAATTTGGAC	1380
Db	1410	AGAGTCTCATTTGGCACTTATCTATGACAAATTTCTGTGCATTTGCTAGGTTAATTTGGAC	1469
OY	1381	AAAACATGTAGCAGGTGTATGATGATTTAGAGTCAAGAACTTAGCATATATAGCTCAAGCT	1440
Db	1470	AAAACATGTAGCAGGTGTATGATGATTTAGAGTCAAGAACTTAGCATATATAGCTCAAGCT	1529

QY	1441	CCCGCTGAGATATGATGATCTCCCTCCCAAGAAAAAAGAAAGGAAAACACCGATTGTGGGCT	1500
Db	1530	CCCGCTGAGATATGATGATCTCCCTCCCAAGAAAAAAGAAAGGAAAACACCGATTGTGGGCT	1589
QY	1501	GCAACATGCGAAGAAAGATACAGCTGAAAAAGAGCGGCTCCTCTAACATGTTTACAACTAT	1560
Db	1590	GCAACATGCGAAGAAAGATACAGCTGAAAAAGAGCGGCTCCTCTAACATGTTTACAACTAT	1649
QY	1561	CAACCCCTGTATCATCCACGCGACGCTTGTGACATTTGTCGCTTGTGTATAGCAAA	1620
Db	1650	CAACCCCTGTATCATCCACGCGACGCTTGTGACATTTGTCGCTTGTGTATAGCAAA	1709
QY	1621	AATTTTGTGAAAAAGTTTGTGCAATGTATGTACAGAGTCAAAAACCGCTTCCGGAGTGC	1680
Db	1710	AATTTTGTGAAAAAGTTTGTGCAATGTATGTACAGAGTCAAAAACCGCTTCCGGAGTGC	1769
QY	1681	CGCTGCAAAAGCAGATGCAACACCAAGCAGTGCCTCGCTTACCTGTGCTGTCCAGATGT	1740
Db	1770	CGCTGCAAAAGCAGATGCAACACCAAGCAGTGCCTCGCTTACCTGTGCTGTCCAGATGT	1829
QY	1741	GACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGGACAGTAAAAATGTGTCC	1800
Db	1830	GACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGGACAGTAAAAATGTGTCC	1889
QY	1801	TGCAAGAAATCTGCAGTATTTCAGCGGGGGTCCAAAAAGATCTATTTGCTGGGACACATCGAC	1860
Db	1890	TGCAAGAAATCTGCAGTATTTCAGCGGGGGTCCAAAAAGATCTATTTGCTGGGACACATCGAC	1949
QY	1861	GTGGCAGGCTGGGGGATTTTTTATCAAGAATCTGTGCAAGAAAAATGAATTCATCTCAGAA	1920
Db	1950	GTGGCAGGCTGGGGGATTTTTTATCAAGAATCTGTGCAAGAAAAATGAATTCATCTCAGAA	2009
QY	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGGAAAGTGTATGATAAA	1980
Db	2010	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGGAAAGTGTATGATAAA	2069
QY	1981	TACATGTGACGCTTCTGTTTCAACTTGAACATATTTTGTGTGTGATGACACCCGGAAG	2040
Db	2070	TACATGTGACGCTTCTGTTTCAACTTGAACATATTTTGTGTGTGATGACACCCGGAAG	2129
QY	2041	GSTAAACAATTTGTTTTGGCAAAATCATTCGGTAAATCAAACTGCTATGCAAAAGTTATG	2100
Db	2130	GSTAAACAATTTGTTTTGGCAAAATCATTCGGTAAATCAAACTGCTATGCAAAAGTTATG	2189
QY	2101	ATGGTTAACGCTGATCAACAGATAGGTATTTTTTSCAAAGAGGCCATTCAGACTGGCGAA	2160
Db	2190	ATGGTTAACGCTGATCAACAGATAGGTATTTTTTSCAAAGAGGCCATTCAGACTGGCGAA	2249
QY	2161	GAGCTGTTTTTGTATACAGATACACGACGCTGATCCTCTGAAGTATGTTCGGCATGAAA	2220
Db	2250	GAGCTGTTTTTGTATACAGATACACGACGCTGATCCTCTGAAGTATGTTCGGCATGAAA	2309
QY	2221	AGAGAAATGAGAAATCCCTTGA 2241	
Db	2310	AGAGAAATGAGAAATCCCTTGA 2310	
RESULT 3			
ADT90243			
ID	ADT90243	standard; DNA; 2600 BP.	
AC	ADT90243;		
XX			
DT	13-JAN-2005	(first entry)	
XX			
DE	Human gene over-expressed in a cancerous cell, SEQ ID 356.		
XX			
KW	cancer; detection; pre-cancerous; antineoplastic; cancer therapy;		
XX	cytostatic; gene therapy; gene; ds.		
OS	Homo sapiens.		
XX			
PN	WO2004091548-A2.		



Db	1590	GCACACTGCAGAAAAGATACACTGAAAAAAGACGGCTCCTCTTAACAGATTGTACACAT	1693
Qy	1561	CAACCTGTGATCATCCACGGCAGCTTGTGAAGTTGTCGCCCTTGTGTATGACAA	1620
Db	1550	CAACCTGTGATCATCCACGGCAGCTTGTGAAGTTGTCGCCCTTGTGTATGACAA	1709
Qy	1621	AAATTTTGTGAAAAGTTTTGTCAATGTATGTCAAGTGTCAAAAACCGCTTCCGGGAATG	1680
Db	1710	AAATTTTGTGAAAAGTTTTGTCAATGTATGTCAAGTGTCAAAAACCGCTTCCGGGAATG	1769
Qy	1681	CGCTGCAAAAGACAGTGCACAACCAAGCAGTGCCTGTACTACTGCTGTCCGAGATGT	1740
Db	1770	CGCTGCAAAAGACAGTGCACAACCAAGCAGTGCCTGTACTACTGCTGTCCGAGATGT	1829
Qy	1741	GACCCCTGACCTCTGTCTTACTTGTGTGAACCGCTGACCAATTGGAACAGTAATAATGTGTC	1800
Db	1830	GACCCCTGACCTCTGTCTTACTTGTGTGAACCGCTGACCAATTGGAACAGTAATAATGTGTC	1889
Qy	1801	TGCAGAAACTGCAGATTCACGCGGGGCTCCAAAAGCATCTATTGCTGGCAACATCTGAC	1866
Db	1890	TGCAGAAACTGCAGATTCACGCGGGGCTCCAAAAGCATCTATTGCTGGCAACATCTGAC	1949
Qy	1861	GTGGCAGGCTGGGGGATTTTTATCAAGAATCCTGTGCAGAAAATGAATTCATCTCAGAA	1920
Db	1950	GTGGCAGGCTGGGGGATTTTTATCAAGAATCCTGTGTGCAGAAAATGAATTCATCTCAGAA	2009
Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAAGGAAAGTGTATGATAA	1980
Db	2010	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAAGGAAAGTGTATGATAA	2069
Qy	1981	TACATGTGCAGCTTTCTGTTCAACTTGAACAATGATTTTGTGTGTGATGCAACCGCAG	2040
Db	2070	TACATGTGCAGCTTTCTGTTCAACTTGAACAATGATTTTGTGTGTGATGCAACCGCAG	2129
Qy	2041	GGTAAACAATAATCGTTTGGCAAAATCAATCCGGTAATTCAAACTGCTATGCAAAAGTTATG	2100
Db	2130	GGTAAACAATAATCGTTTGGCAAAATCAATCCGGTAATTCAAACTGCTATGCAAAAGTTATG	2189
Qy	2101	ATGTGTAAACGGTGATCACAGATAGGTATTTTGTGCCAAGAGCCATCCAGACTGGCGAA	2166
Db	2190	ATGTGTAAACGGTGATCACAGATAGGTATTTTGTGCCAAGAGCCATCCAGACTGGCGAA	2249
Qy	2161	GAGCTGTTTTTTGATTACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGCATCGAA	2220
Db	2250	GAGCTGTTTTTTGATTACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGCATCGAA	2309
Qy	2221	AGAGAAATGGAATCCCTTGA	2241
Db	2310	AGAGAAATGGAATCCCTTGA	2330
RESULT 4			
ADT89975			
ID	ADT89975	standard; DNA; 2600 BP.	
XX	ADT89975;		
AC			
XX			
DT	13-JAN-2005	(first entry)	
XX			
DE	Human gene over-expressed in a cancerous cell, SEQ ID 88.		
XX			
KW	cancer; detection; pre-cancerous; antineoplastic; cancer therapy;		
KW	cytostatic; gene therapy; gene; de.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2004091548-A2.		
XX			
PD	28-OCT-2004.		
XX			
PF	15-APR-2004; 2004WO-US009289.		
XX			
PR	15-APR-2003; 2003US-0462895P.		

XX	(AVAL-) AVALON PHARM INC.
PA	Strovel JW, Cain CB, Horrigan SK, Augustus M;
XX	WPI; 2004-784489/77.
XX	
PT	Diagnosing cancer in a mammal comprises determining amplification of
XX	a specific gene, e.g., serine/threonine kinase 6, in the genome of a
PS	mammal.
XX	
PS	Claim 3; SEQ ID NO 88; 92bp; English.
CC	
XX	The invention relates to a novel method for diagnosing cancer in a
CC	mammal. The method comprises determining amplification of specific genes,
CC	given in the specification, in the genome of a mammal. The invention
CC	further comprises methods for: diagnosing cancer or a pre-cancerous
CC	condition in a mammal; inhibiting cancer or a pre-cancerous condition in
CC	a mammalian cell; identifying an agent having therapeutic activity in a
CC	human patient; identifying an anti-neoplastic agent; determining the
CC	cancerous status of a cell; identifying a compound as an anti-neoplastic
CC	agent; treating cancer; monitoring the progress of cancer therapy in a
CC	patient; determining the likelihood of success of cancer therapy in a
CC	patient; producing test data with respect to the anti-neoplastic activity
CC	of a compound; and determining the progress of a treatment for cancer in
CC	a patient following commencement of a cancer treatment on the patient.
CC	The methods and compositions of the invention have cytostatic activity
CC	and may be used in gene therapy. The method is useful in diagnosing or
CC	treating cancer in a mammal. This polynucleotide sequence represents a
CC	gene over-expressed in a cancerous cell, used in the novel cancer
CC	diagnosis method of the invention.
XX	
SQ	Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;
Query Match	100.0%; Score 2241; DB 13; Length 2600;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
DB	
QY	1 ATGGGCCAGACTGCGGAAGAATCTTGAAAGGACCAAGTTTGTGGCGGAGCGTATAAA 60
DB	90 ATGGGCCAGACTGCGGAAGAATCTTGAAAGGACCAAGTTTGTGGCGGAGCGTATAAA 149
QY	61 TCAGAGTCACAGCCGACGTAGACAGTCAAGAGGTTCAACAGCGCATGAAAGTAAGTAGT 120
DB	150 TCAGAGTCACAGCCGACGTAGACAGTCAAGAGGTTCAACAGCGCATGAAAGTAAGTAGT 209
QY	121 ATGTTAAGTCCAAACGTCAGAAAATTTTGGAAAAGAAAGSAAATCTTAAACCAGAAATGG 180
DB	210 ATGTTAAGTCCAAACGTCAGAAAATTTTGGAAAAGAAAGSAAATCTTAAACCAGAAATGG 269
QY	181 AAACAGCGAAGATACAGCTGTGTGCACATCTGACTTGTGTAGCTCATTGCGCGGACT 240
DB	270 AAACAGCGAAGATACAGCTGTGTGCACATCTGACTTGTGTAGCTCATTGCGCGGACT 329
QY	241 AGGAGGTGTTCGCTGACCAAGTGAATTTTCCAACAAGATATCCATTAAAGACT 300
DB	330 AGGAGGTGTTCGCTGACCAAGTGAATTTTCCAACAAGATATCCATTAAAGACT 389
QY	301 CTGAATGACGTTGCTTCAGTACCCATAANGATTTCTTGCTCCCCCTACAGCAGAAATTT 360
DB	390 CTGAATGACGTTGCTTCAGTACCCATAANGATTTCTTGCTCCCCCTACAGCAGAAATTT 449
QY	361 ATGTGTGAAGATGAACCTGTTTACATACATCTCTTANATGGAGATGAAGTTTTAGAT 420
DB	450 ATGTGTGAAGATGAACCTGTTTACATACATCTCTTANATGGAGATGAAGTTTTAGAT 509
QY	421 CAGGATGTTATCTTTCAATTGAGAACTAATTAATAAATTTATGATGGGAAAGTACACGGGAT 480
DB	510 CAGGATGTTATCTTTCAATTGAGAACTAATTAATAAATTTATGATGGGAAAGTACACGGGAT 569
QY	481 AGAGATGTGGGTTTATTAATGATGAATAAATTTTGGAGATTGGTGAATGCCCTTGATCA 540
DB	570 AGAGATGTGGGTTTATTAATGATGAATAAATTTTGGAGATTGGTGAATGCCCTTGATCA 629

[illegible]

QY	1621	AAATTTTGTGAAAAGTTTGTGCAATGATGTGAGAGGTGCAAAAACGGCTTTCGGGAGTGC	1680
Db	1710	AAATTTTGTGAAAAGTTTGTGCAATGATGTGAGAGGTGCAAAAACGGCTTTCGGGAGTGC	1769
QY	1681	CGCTGCAGGAAGACAGTGCACAACAACGACAGTGCCTGCTACCTGCTGTCCAGAGTGT	1740
Db	1770	CGCTGCAGGAAGACAGTGCACAACAACGACAGTGCCTGCTACCTGCTGTCCAGAGTGT	1829
QY	1741	GACCCCTGACCTGTGTCTTACTTGTGTGAGCCGCTGACCAATTGGGACAGTAAATAATGTGTCC	1800
Db	1830	GACCCCTGACCTGTGTCTTACTTGTGTGAGCCGCTGACCAATTGGGACAGTAAATAATGTGTCC	1889
QY	1801	TGCAAGAACTGACAGATATTCACGCGGGGCTCCAAAAAGCATTTATTTGCTGGACCATCTGCAC	1860
Db	1890	TGCAAGAACTGACAGATATTCACGCGGGGCTCCAAAAAGCATTTATTTGCTGGACCATCTGCAC	1949
QY	1861	GTGGCAGGCTGGGGGGAATTTTATCAAAAGATCCTGTGCAGAAAAATGAATTCACTGCAGAA	1920
Db	1950	GTGGCAGGCTGGGGGGAATTTTATCAAAAGATCCTGTGCAGAAAAATGAATTCACTGCAGAA	2009
QY	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGCAGAAAGAGGAAAAGTGTATGATPAA	1980
Db	2010	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGCAGAAAGAGGAAAAGTGTATGATPAA	2069
QY	1981	TACATGTGCAGCTTCTGTTTCACTTGAACAATGATTTTGTGTGTGATTCGACCCGCGAC	2040
Db	2070	TACATGTGCAGCTTCTGTTTCACTTGAACAATGATTTTGTGTGTGATTCGACCCGCGAC	2129
QY	2041	GGTAACAAAATTCGTTTGTGCAATCATTCGGTAAATCCAACTGCTATGCAAAAGTTATG	2100
Db	2130	GGTAACAAAATTCGTTTGTGCAATCATTCGGTAAATCCAACTGCTATGCAAAAGTTATG	2189
QY	2101	ATGGTTAAACGGTGTATCAACAGATAGTATTTTGTGCCAAGAGAGCCATCCAGACTGGCGAA	2160
Db	2190	ATGGTTAAACGGTGTATCAACAGATAGTATTTTGTGCCAAGAGAGCCATCCAGACTGGCGAA	2249
QY	2161	GAGCTGTTTTTGTGATTAACAATACACGACGAGCTGATGCCCTGAAGTATGTGCGCATGAA	2220
Db	2250	GAGCTGTTTTTGTGATTAACAATACACGAGCTGATGCCCTGAAGTATGTGCGCATGAA	2309
QY	2221	AGAGAAATGGAATCCCTTGA 2241	
Db	2310	AGAGAAATGGAATCCCTTGA 2330	
RESULT 5			
ADT90094			
ID	ADT90094	standard; DNA; 2600 BP.	
AC	ADT90094;		
XX			
DT	13-JAN-2005	(first entry)	
XX			
DE	Human gene over-expressed in a cancerous cell, SEQ ID 207.		
XX			
KN	cancer; detection; pre-cancerous; antineoplastic; cancer therapy;		
XX			
KW	cytostatic; gene therapy; gene; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2004091548-A2.		
XX			
PD	28-OCT-2004.		
XX			
PF	15-APR-2004; 2004WO-US009289.		
XX			
PR	15-APR-2003; 2003US-0462895P.		
XX			
PA	(AVAL- ) AVALON PHARM INC.		
XX			
PI	Strovel JM, Cain CB, Horrigan SK, Augustus M;		
XX	WPI; 2004-784489/77.		



XX Diagnosing cancer in a mammal comprises determining amplification of  
PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
PT mammal.  
PS Claim 3; SEQ ID NO 207; 92pp; English.  
XX  
CC The invention relates to a novel method for diagnosing cancer in a  
CC mammal. The method comprises determining amplification of specific genes,  
CC given in the specification, in the genome of a mammal. The invention  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
CC human patient; identifying an anti-neoplastic agent; determining the  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytostatic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;  
Query Match 100.0%; Score 2241; DB 13; Length 2600;  
Beeb Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGAGACGTGTAAAA 60  
DB 90 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGAGACGTGTAAAA 149  
QY 61 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGACGAGTGAAGTAAGAGT 120  
DB 150 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGACGAGTGAAGTAAGAGT 209  
QY 121 ATGTTTATTTCCAAATGTCAGAAATTTTGGAAAGAGGAATTTTAAACCAAGAAATG 180  
DB 210 ATGTTTATTTCCAAATGTCAGAAATTTTGGAAAGAGGAATTTTAAACCAAGAAATG 269  
QY 181 AAACGACGAGAGATACAGCTGTGACATCTGACTTGTGAGCTCATTTGGCGGAGCT 240  
DB 270 AAACGACGAGAGATACAGCTGTGACATCTGACTTGTGAGCTCATTTGGCGGAGCT 329  
QY 241 AGGAGTGTTCGGTGAACAAGTATTTTCCAAACAAGTATCTCCATTAAAGCT 300  
DB 330 AGGAGTGTTCGGTGAACAAGTATTTTCCAAACAAGTATCTCCATTAAAGCT 389  
QY 301 CTGAATGAGTGTTCCTCAATCCCATANATTTCTGTGTCCTCCCTCAGAGAAATTT 360  
DB 390 CTGAATGAGTGTTCCTCAATCCCATANATTTCTGTGTCCTCCCTCAGAGAAATTT 449  
QY 361 ATGTGGAAGATGAATCTGTTTACATAACATTCCTTATATGGAGATGAAGTTTATAG 420  
DB 450 ATGTGGAAGATGAATCTGTTTACATAACATTCCTTATATGGAGATGAAGTTTATAG 509  
QY 421 CAGATGCTACTTTTATGAAAGCTAATAAAAATTTATGATGGAAAGTACACGGGAT 480  
DB 510 CAGATGCTACTTTTATGAAAGCTAATAAAAATTTATGATGGAAAGTACACGGGAT 569  
QY 481 AGAGAAATGTGGCTTATATAAGTAAATTTTGTGGAGTGTGTGAATGCCCTTGTGTA 540  
DB 570 AGAGAAATGTGGCTTATATAAGTAAATTTTGTGGAGTGTGTGAATGCCCTTGTGTA 629  
QY 541 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
DB 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
QY 601 AAAGATCTGAGAGATCACCGAGATGATAAGAAAGCCGCCACCTCGGAAATTTCTTCT 660

DB 690 AAAGATCTGAGAGATCACCGAGATGATAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
QY 661 GATTAATTTTGAAGCATTTTCTCAATGTTTTCAGATTAAGGACACGAGAGAACTA 720  
DB 750 GATTAATTTTGAAGCATTTTCTCAATGTTTTCAGATTAAGGACACGAGAGAACTA 809  
QY 721 AAGGAAAAATATTAAGAACTACACGAAACAGAGCTCCAGCGCATTTCTCTGATATG 780  
DB 810 AAGGAAAAATATTAAGAACTACACGAAACAGAGCTCCAGCGCATTTCTCTGATATG 869  
QY 781 ACCCCGAAATATTAAGAACTACACGAAATGCTTAATCTGTTCAAGAGACAAAGTTTCACTCC 840  
DB 870 ACCCCGAAATATTAAGAACTACACGAAATGCTTAATCTGTTCAAGAGACAAAGTTTCACTCC 929  
QY 841 TTTTCATAGCTTTTCTGTAGCGATGTTTAAATATGATGCTTCTTACATCTTTTCAT 900  
DB 930 TTTTCATAGCTTTTCTGTAGCGATGTTTAAATATGATGCTTCTTACATCTTTTCAT 989  
QY 901 GCAACACCCCAACTTATTAAGCGGAGAGACACAGAAACAGCTTTAGACAAACCAATCTTGT 960  
DB 990 GCAACACCCCAACTTATTAAGCGGAGAGACACAGAAACAGCTTTAGACAAACCAATCTTGT 1049  
QY 961 GCAACACGTTTACACGCAATTTTGAAGGAGCAAGAGTTTGTGCTGCTCACCGCT 1020  
DB 1050 GCAACACGTTTACACGCAATTTTGAAGGAGCAAGAGTTTGTGCTGCTCACCGCT 1109  
QY 1021 GAGCGATTAAGACCCCAACCAACGTCAGAGGCGCGAGAAAGAGACGCTTCCCAAT 1080  
DB 1110 GAGCGATTAAGACCCCAACCAACGTCAGAGGCGCGAGAAAGAGACGCTTCCCAAT 1169  
QY 1081 AACATGACAGGCGCCAGACACCCCAACCAATTAATGCTGGAATTAAGATTAAGACAGT 1140  
DB 1170 AACATGACAGGCGCCAGACACCCCAACCAATTAATGCTGGAATTAAGATTAAGACAGT 1229  
QY 1141 GATTAAGAGAGAGGAGGAGTGAACGAGGAGAGAGAAATGATTAAGAAAGAGAGAG 1200  
DB 1230 GATTAAGAGAGAGGAGGAGTGAACGAGGAGAGAGAAATGATTAAGAAAGAGAGAG 1289  
QY 1201 AAAGATGAATCTTGAAGCTCTCTGAAGCAATTTCTGCTGTCAACCAACCAATTAAGATG 1260  
DB 1290 AAAGATGAATCTTGAAGCTCTCTGAAGCAATTTCTGCTGTCAACCAACCAATTAAGATG 1349  
QY 1261 AAGCAAAATATTAAGCTCTGAGATGATGAGATGATGATGATGATGATGATGATGATG 1320  
DB 1350 AAGCAAAATATTAAGCTCTGAGATGATGAGATGATGATGATGATGATGATGATGATG 1409  
QY 1321 AGAGTCCCATTTGCACTTACTATGACAAATTTCTGTGCAATTTGCTAGGTTAATGGAGC 1380  
DB 1410 AGAGTCCCATTTGCACTTACTATGACAAATTTCTGTGCAATTTGCTAGGTTAATGGAGC 1469  
QY 1381 AAAACATGTAGACAGTGTATGAGTTTGAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1440  
DB 1470 AAAACATGTAGACAGTGTATGAGTTTGAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1529  
QY 1441 CCCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
DB 1530 CCCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1589  
QY 1501 GCACATGCAAGAAAGATACAGCTGAAAAAGGACGGCTCTCTTAACCAATGTTTACAACTAT 1560  
DB 1590 GCACATGCAAGAAAGATACAGCTGAAAAAGGACGGCTCTCTTAACCAATGTTTACAACTAT 1649  
QY 1561 CAACCTGTGATCATCAACGACGCTGTGTGACAGTTCTGTGCTGTGTGTATGACAA 1620  
DB 1650 CAACCTGTGATCATCAACGACGCTGTGTGACAGTTCTGTGCTGTGTGTATGACAA 1709  
QY 1621 AATTTTGTGAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATG 1680  
DB 1710 AATTTTGTGAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATG 1769  
QY 1681 CGCTGCAAGACAGTGCACACAGAGAGGCGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1740



DB 1770 CGCTCAAGACAGTGCACCAACCAAGCAGTCCCGTGTACTGCTGCCAGAGTGT 1829  
CC  
CC 1741 GACCTGACCTCTGTCTTACTTGTGAGACCGCTGACCAATGGGACAGTAAATGTGTC 1800  
CC  
DB 1830 GACCTGACCTCTGTCTTACTTGTGAGACCGCTGACCAATGGGACAGTAAATGTGTC 1889  
CC  
QY 1801 TGCAGAACTGACAGTATTCAGCGGGCTCCAAAAGCATCTATTCTGGCACCATTCTGAC 1860  
CC  
DB 1890 TGCAGAACTGACAGTATTCAGCGGGCTCCAAAAGCATCTATTCTGGCACCATTCTGAC 1949  
CC  
QY 1861 GTGGCAGCTGGGGGATTTTATTCAAAGATCTGTGACAGAAAATGATTCATCTCAGA 1920  
CC  
DB 1950 GTGGCAGCTGGGGGATTTTATTCAAAGATCTGTGACAGAAAATGATTCATCTCAGA 2009  
CC  
QY 1921 TACTGTGAGAGATTTATTTCTCAAGTAAAGCTGACAGAAAGGGAAGTGTATGATAA 1980  
CC  
DB 2010 TACTGTGAGAGATTTATTTCTCAAGTAAAGCTGACAGAAAGGGAAGTGTATGATAA 2069  
CC  
QY 1981 TACATGTGACGCTTCTGTTCGAATCTGAAACATGATTTGTGTGATGCAACCCGCAAG 2040  
CC  
DB 2070 TACATGTGACGCTTCTGTTCGAATCTGAAACATGATTTGTGTGATGCAACCCGCAAG 2129  
CC  
QY 2041 GGTAAACAAATTCGTTTGCATATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2100  
CC  
DB 2130 GGTAAACAAATTCGTTTGCATATCATTCGTTAAATCCAACTGCTATGCAAAAGTTATG 2189  
CC  
QY 2101 ATGTTAAACGTTGATCAGAGTAGGTATTTTGGCAGAGAGCCATCCAGCTGCGCA 2160  
CC  
DB 2190 ATGTTAAACGTTGATCAGAGTAGGTATTTTGGCAGAGAGCCATCCAGCTGCGCA 2249  
CC  
QY 2161 GAGCTGTTTGTGATTAAGATACAGACGAGCGAGTGAAGCCCTGAAGTATGGGATCGAA 2220  
CC  
DB 2250 GAGCTGTTTGTGATTAAGATACAGACGAGCGAGTGAAGCCCTGAAGTATGGGATCGAA 2309  
CC  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
CC  
DB 2310 AGAGAAATGAAATCCCTTGA 2330  
CC  
RESULT 6  
ADT90198  
ID ADT90198 standard; DNA; 2600 BP.  
XX  
AC ADT90198;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE Human gene over-expressed in a cancerous cell, SEQ ID 311.  
XX  
KM cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
XX  
KM cytostatic; gene therapy; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2004091548-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 15-APR-2004; 2004WO-US009289.  
XX  
PR 15-APR-2003; 2003US-0462895P.  
XX  
PA (AVAL-) AVALON PHARM INC.  
XX  
PI Stovel JW, Cain CB, Horrigan SK, Augustus M;  
XX  
DR WPI; 2004-784489/77.  
XX  
PT Diagnosing cancer in a mammal comprises determining amplification of  
PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
XX  
XX mammal.  
PS Claim 3; SEQ ID NO 311; 92pp; English.

XX  
CC The invention relates to a novel method for diagnosing cancer in a  
CC mammal. The method comprises determining amplification of specific genes,  
CC given in the specification, in the genome of a mammal. The invention  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
CC human patient; identifying an anti-neoplastic agent; determining the  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytostatic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;  
Query Match 100.0%; Score 2241; DB 13; Length 2600;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGCCGACCTGGGAAAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAGCTGTAA 60  
DB 90 ATGGGCCGACCTGGGAAAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAGCTGTAA 149  
DB 61 TCAGAGTACATCCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAAGTAAAGT 120  
DB 150 TCAGAGTACATCCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAAGTAAAGT 209  
QY 121 ATGTTAGTCCATTCGTCAAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATGG 180  
DB 210 ATGTTAGTCCATTCGTCAAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATGG 269  
QY 181 AAACAGCGAAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGCGGAGCT 240  
DB 270 AAACAGCGAAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGCGGAGCT 329  
QY 241 AGGAGTGTGGGAGACAGTGAATTTTCCACACAAGTATCCCATTAAGACT 300  
DB 330 AGGAGTGTGGGAGACAGTGAATTTTCCACACAAGTATCCCATTAAGACT 389  
QY 301 CTGAATGCAATGCTTCACTACCAATATGATCTTGTCTCCCTCAAGCAATTTT 360  
DB 390 CTGAATGCAATGCTTCACTACCAATATGATCTTGTCTCCCTCAAGCAATTTT 449  
QY 361 ATGTGGAAGATGAATGCTTTTACATTAACATCTTATATGAGGAGATGAAGTTTAAAT 420  
DB 450 ATGTGGAAGATGAATGCTTTTACATTAACATCTTATATGAGGAGATGAAGTTTAAAT 509  
QY 421 CAGATGTGACTTTCATTAAGAACTAATAAATAATTAAGTGAAGGAAAGTCAAGGGAGT 480  
DB 510 CAGATGTGACTTTCATTAAGAACTAATAAATAATTAAGTGAAGGAAAGTCAAGGGAGT 569  
QY 481 AGAGATGTGGGTTTAAATGAAGAAATTTTGTGAGTGTGATGCTTGTGTCA 540  
DB 570 AGAGATGTGGGTTTAAATGAAGAAATTTTGTGAGTGTGATGCTTGTGTCA 629  
QY 541 TATTAATGATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 630 TATTAATGATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 601 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCACTCGGAAATTTCTCT 660  
DB 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCACTCGGAAATTTCTCT 749  
QY 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGAGACAGCAAGAACTA 720  
DB 750 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGAGACAGCAAGAACTA 809

Qy 721 AAGAAAAATATAAGAACTCAACGGAACAGAGCTCCAGGCGCACTTCTCTGAATGT 780  
 Db 810 AAGAAAAATATAAGAACTCAACGGAACAGAGCTCCAGGCGCACTTCTCTGAATGT 869  
 Qy 781 ACCCCCAACATAGATGAGCCAAATGCTAAATCTGTTCAGAGAGAGCAAGCTTACATCC 840  
 Db 870 ACCCCCAACATAGATGAGCCAAATGCTAAATCTGTTCAGAGAGAGCAAGCTTACATCC 929  
 Qy 841 TTTCAATAGCTTTTCTGTAGGCGAATGTTTAAATATGATCTGCTTCAATCTCTTTCAT 900  
 Db 930 TTTCAATAGCTTTTCTGTAGGCGAATGTTTAAATATGATCTGCTTCAATCTCTTTCAT 989  
 Qy 901 GCAACACCCCAACATTAATAGCGGAGAAACAAGAAACAGCTTATAGCAACAACTTGT 960  
 Db 990 GCAACACCCCAACATTAATAGCGGAGAAACAAGAAACAGCTTATAGCAACAACTTGT 1049  
 Qy 961 GGACCAACAGTGTATACAGCAATTTGGAGGAGCAAGAGTTTGTCTGCTCTCAACGCT 1020  
 Db 1050 GGACCAACAGTGTATACAGCAATTTGGAGGAGCAAGAGTTTGTCTGCTCTCAACGCT 1109  
 Qy 1021 GAGCGGATTAAGAACCCCAACAAAGCTCAGAGAGGCGCAGAGAGAGCGCTTCCCAAT 1080  
 Db 1110 GAGCGGATTAAGAACCCCAACAAAGCTCAGAGAGGCGCAGAGAGAGCGCTTCCCAAT 1169  
 Qy 1081 AACAGTAGCAGGCGCCAGCAACCCCAACATTAATGTCTGAATCAAGAGATACAGACAT 1140  
 Db 1170 AACAGTAGCAGGCGCCAGCAACCCCAACATTAATGTCTGAATCAAGAGATACAGACAT 1229  
 Qy 1141 GATAGGGAAGCAAGGAGCTGAAACGGGGGGAAGAAACATGATTAAGAAAGAAAGAGAG 1200  
 Db 1230 GATAGGGAAGCAAGGAGCTGAAACGGGGGGAAGAAACATGATTAAGAAAGAAAGAGAG 1289  
 Qy 1201 AAAGATGAACCTTGAAGCTCTCTGAAGCAAAATCTCGGTGTCAAAACACATTAAGATG 1260  
 Db 1290 AAAGATGAACCTTGAAGCTCTCTGAAGCAAAATCTCGGTGTCAAAACACATTAAGATG 1349  
 Qy 1261 AAGCCAAATATTGAACCTCTCTGAAGATGTGAGTGTGTCTGAAGCTCAATGTTT 1320  
 Db 1350 AAGCCAAATATTGAACCTCTCTGAAGATGTGAGTGTGTCTGAAGCTCAATGTTT 1409  
 Qy 1321 AGAGTCTCTATTTGGAATTAATTAAGCAATTTCTGTGCAATTTGCTTAAGTTTATGGACC 1380  
 Db 1410 AGAGTCTCTATTTGGAATTAATTAAGCAATTTCTGTGCAATTTGCTTAAGTTTATGGACC 1469  
 Qy 1381 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATATAGCTCCAGCT 1440  
 Db 1470 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATATAGCTCCAGCT 1529  
 Qy 1441 CCCGCTGAGATGTGATATCTCTCTCAAGAAAAAGAAAGAGAAACACCGGTTGTGGCT 1500  
 Db 1530 CCCGCTGAGATGTGATATCTCTCTCAAGAAAAAGAAAGAGAAACACCGGTTGTGGCT 1589  
 Qy 1501 GCACTCTGCAAGAAATACAGCTGAAAAAGACGGCTCTCTTAACATATTTTCAACTAT 1560  
 Db 1590 GCACTCTGCAAGAAATACAGCTGAAAAAGACGGCTCTCTTAACATATTTTCAACTAT 1649  
 Qy 1561 CAACCTGTGATCATCAAGGAGCGCTGTGACAGTGTGTCGCCCTGTGTGTATAGACAA 1620  
 Db 1650 CAACCTGTGATCATCAAGGAGCGCTGTGACAGTGTGTCGCCCTGTGTGTATAGACAA 1709  
 Qy 1621 AATTTTGTGAAAAAGTTTGTCAATGTATGTCAAGTGTCAAAACCGCTTTCGGGATGC 1680  
 Db 1710 AATTTTGTGAAAAAGTTTGTCAATGTATGTCAAGTGTCAAAACCGCTTTCGGGATGC 1769  
 Qy 1681 CGCTGCAAGCAACAGTGTCAACCAAGAGTGCCTGTCTACCTGTGTGTGTGTGTGTGT 1740  
 Db 1770 CGCTGCAAGCAACAGTGTCAACCAAGAGTGCCTGTCTACCTGTGTGTGTGTGTGTGT 1829  
 Qy 1741 GACCCCTGACTCTGTCTTACTTGTGTGAGCGCTGTGACCAATGTGGAACATTAATGTGCC 1800  
 Db 1830 GACCCCTGACTCTGTCTTACTTGTGTGAGCGCTGTGACCAATGTGGAACATTAATGTGCC 1889

Qy 1801 TGCAGAACTGCAAGTATTCAGCGGGGCTCCAAAAAGCATTTATGCTGCGCACTTGTAC 1860  
 Db 1890 TGCAGAACTGCAAGTATTCAGCGGGGCTCCAAAAAGCATTTATGCTGCGCACTTGTAC 1949  
 Qy 1861 GTGGCAGGCTGGGGGATTTTATTAACAAGATCTGTGCAAGAAAAATGAATTCATCTGAAA 1920  
 Db 1950 GTGGCAGGCTGGGGGATTTTATTAACAAGATCTGTGCAAGAAAAATGAATTCATCTGAAA 2009  
 Qy 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAAAGTATGATAAA 1980  
 Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAAAGTATGATAAA 2069  
 Qy 1981 TACATGTGACAGCTTCTGTTCACACTTAACAAGATTTTGTGTGTGTGTGTGTGTGTGT 2040  
 Db 2070 TACATGTGACAGCTTCTGTTCACACTTAACAAGATTTTGTGTGTGTGTGTGTGTGTGT 2129  
 Qy 2041 GGTACAAAAATTTGTTTGAATTAATCATTTGGTAAATCCAAACTGCTATGCAAAAGTTATG 2100  
 Db 2130 GGTACAAAAATTTGTTTGAATTAATCATTTGGTAAATCCAAACTGCTATGCAAAAGTTATG 2189  
 Qy 2101 ATGTTAACGATGATACAGAGATAGTATTTTGTCCAAAGAGAGCCATCCAGACTGGCGAA 2160  
 Db 2190 ATGTTAACGATGATACAGAGATAGTATTTTGTCCAAAGAGAGCCATCCAGACTGGCGAA 2249  
 Qy 2161 GAGCTGTTTTTTGATTAACAATACAGCCAGGCTGATGCTCTGAAGTATGTCTGGATGAAA 2220  
 Db 2250 GAGCTGTTTTTTGATTAACAATACAGCCAGGCTGATGCTCTGAAGTATGTCTGGATGAAA 2309  
 Qy 2221 AGAGAAATGGAATCCCTTGA 2241  
 Db 2310 AGAGAAATGGAATCCCTTGA 2330

RESULT 7  
 ADQ22638  
 ID ADQ22638 standard; DNA; 2711 BP.  
 XX AC ADQ22638;  
 XX AC 26-AUG-2004 (first entry)  
 XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5458.  
 XX KW soft tissue sarcoma; cytoslastic; gene therapy; vaccine; screening; human;  
 XX KW ds.  
 XX OS Homo sapiens.  
 XX OS WO2004048938-A2.  
 XX PD 10-JUN-2004.  
 XX PD 26-NOV-2003; 2003WO-US038193.  
 XX PR 26-NOV-2002; 2002US-0429739P.  
 XX PR 26-NOV-2002; 2002US-0429739P.  
 XX PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX PA Aziz N, Ginsburg WM, Zlotnik A;  
 XX PI WPI; 2004-441208/41.  
 XX DR  
 XX XX Early detection of soft tissue sarcoma comprises determining expression  
 XX PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 XX PT and comparing the gene expression, also useful in treating soft tissue  
 XX PT sarcoma.  
 XX PS Example 2; SEQ ID NO 5458; 210pp; English.  
 XX CC The invention relates to a novel method for detecting soft tissue sarcoma  
 XX CC which comprises obtaining a first soft tissue sample from an individual  
 XX CC and a normal soft tissue sample from the same or different individual,  
 XX CC determining the expression of a gene in both samples and comparing the

expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

XX Sequence 2711 BP; 865 A; 547 C; 658 G; 641 T; 0 U; 0 Other;

Query Match 100.0%; Score 2241; DB 12; Length 2711;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0; Matches 2241; Conservative 0;

1 ATGGCCAGACTGGGAGAAATCTGAGAGGACAGTTGTGGCGGAGCGTGTAAA 60  
167 ATGGCCAGACTGGGAGAAATCTGAGAGGACAGTTGTGGCGGAGCGTGTAAA 226  
61 TCAGGTACATCGACTGAGACAGCTCAAGAGTTCAAGCAGCTGATGAATAAGAT 120  
227 TCAGGTACATCGACTGAGACAGCTCAAGAGTTCAAGCAGCTGATGAATAAGAT 286  
121 ATGTTAGTTCAATCGTCAAAAATTTTGAAGAAGGGAATCTTAAACAGAAATGG 180  
287 ATGTTAGTTCAATCGTCAAAAATTTTGAAGAAGGGAATCTTAAACAGAAATGG 346  
181 AAACAGCGAAGATACAGCTGTGACATCTGACTTGTGAGCTCATTTGGCGGAGCT 240  
347 AAACAGCGAAGATACAGCTGTGACATCTGACTTGTGAGCTCATTTGGCGGAGCT 406  
241 AGGAGTGTTCGGTGAACAGTGAATTTTCCAAACAGAGTCAATCCATTAAAGCT 300  
407 AGGAGTGTTCGGTGAACAGTGAATTTTCCAAACAGAGTCAATCCATTAAAGCT 466  
301 CTGAATGAGTGTTCGGTGAACAGTGAATTTTCCAAACAGAGTCAATCCATTAAAGCT 360  
467 CTGAATGAGTGTTCGGTGAACAGTGAATTTTCCAAACAGAGTCAATCCATTAAAGCT 526  
361 ATGTGGAAGATGAACCTGTTTACATTAATCTCTTATATGAGAGTGAATTTTACAT 420  
527 ATGTGGAAGATGAACCTGTTTACATTAATCTCTTATATGAGAGTGAATTTTACAT 586  
421 CAGATGCTACTTCAATGAAGAATTAATTAATATGATGAGAAAGTACACGGGAT 480  
587 CAGATGCTACTTCAATGAAGAATTAATTAATATGATGAGAAAGTACACGGGAT 646  
481 AGAGATGTGGTGTATTAATGAATAATTTTGTGAGTGTGAATGCCCTTGCTCAA 540  
647 AGAGATGTGGTGTATTAATGAATAATTTTGTGAGTGTGAATGCCCTTGCTCAA 706  
541 TATATGATGATGAGATGATGATGATGAGAGATCTCTGAAGAAAGAGAAAGACAG 600  
707 TATATGATGATGAGATGATGATGATGAGAGATCTCTGAAGAAAGAGAAAGACAG 766  
601 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCCACTCTGGAATTTCTTCT 660  
767 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCCACTCTGGAATTTCTTCT 826  
661 GATATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGAGACAGCAAGAAACTA 720  
827 GATATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGAGACAGCAAGAAACTA 886  
721 AAGGAAAATATTAAGAACTCAACGACAGCTCCGAGCGCACTTCTCTGAAATGT 780  
887 AAGGAAAATATTAAGAACTCAACGACAGCTCCGAGCGCACTTCTCTGAAATGT 946  
781 ACCCCCAATATGATGAGCAATATGCTAAATCTGTTGAGAGAGACCAAGCTTACATCC 840  
947 ACCCCCAATATGATGAGCAATATGCTAAATCTGTTGAGAGAGACCAAGCTTACATCC 1006  
841 TTTCATAGGCTTTTCTGTAAGCGATGTTTAAATATGATGCTTCTTACATCTTTTCAAT 900

1007 TTTTATAGGCTTTTCTGTAAGCGATGTTTAAATATGATGCTTCTTACATCTTTTCAAT 1066  
901 GCAACACCCACATTAATTAAGGAGAAACACAGAAACAGCTCTGACACCAACTTGT 960  
1067 GCAACACCCACATTAATTAAGGAGAAACACAGAAACAGCTCTGACACCAACTTGT 1126  
961 GGACCAAGTGTATACACATTTTGAAGGAGCAAGAGTTTGTGCTGCTCCACCGCT 1020  
1127 GGACCAAGTGTATACACATTTTGAAGGAGCAAGAGTTTGTGCTGCTCCACCGCT 1186  
1021 GAGCGATTAAGACCCCAACCAAAAGTCCAGAGGCGCCAGAGAGACAGCTTCCCAT 1080  
1187 GAGCGATTAAGACCCCAACCAAAAGTCCAGAGGCGCCAGAGAGAGAGCTTCCCAT 1246  
1081 AACATGACAGGCGCCACACCCCACTTAATGTGCTGGAATCAAGAGATACAGACGT 1140  
1247 AACATGACAGGCGCCACACCCCACTTAATGTGCTGGAATCAAGAGATACAGACGT 1306  
1141 GATTGGAGAGCAGGAGCTGAAGACGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1200  
1307 GATTGGAGAGCAGGAGCTGAAGACGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1366  
1201 AAAGATGAACCTTCAGAGCTCTCTGAACCAAAATCTCGGTGTCAACCAATTAAGATG 1260  
1367 AAAGATGAACCTTCAGAGCTCTCTGAACCAAAATCTCGGTGTCAACCAATTAAGATG 1426  
1261 AAGCCAAATATTAAGACCTCTGAGATGTGAGAGTGTGTCTGAAGCTCAATGTTT 1320  
1427 AAGCCAAATATTAAGACCTCTGAGATGTGAGAGTGTGTGTGAAGCTCAATGTTT 1486  
1321 AGAGTCTCATTTGGACCTTAATGAACAATTTGTGTGCAATGCTTAAGTTAATGGAGC 1380  
1487 AGAGTCTCATTTGGACCTTAATGAACAATTTGTGTGCAATGCTTAAGTTAATGGAGC 1546  
1381 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAGATCTAGCATATAGCTCAGCT 1440  
1547 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAGATCTAGCATATAGCTCAGCT 1606  
1441 CCCGCTGAGATGTGATTAATCTCTCAAGAGAAAGAGAGAAACACCGGTTGGGCT 1500  
1607 CCCGCTGAGATGTGATTAATCTCTCAAGAGAAAGAGAGAAACACCGGTTGGGCT 1666  
1501 GCACTGCAAGAAAGATACAGTGAAGAAAGAGCGCTCTCTTAACATGTTTACAACTAT 1560  
1667 GCACTGCAAGAAAGATACAGTGAAGAAAGAGCGCTCTCTTAACATGTTTACAACTAT 1726  
1561 CAACCTGTGATCATCAAGGAGCGCTGTGACAGTTGCTGCGCTGTGTATAGACACA 1620  
1727 CAACCTGTGATCATCAAGGAGCGCTGTGACAGTTGCTGCGCTGTGTATAGACACA 1786  
1621 AATTTTGTGAAGATTTTGTCAATGTATGATTAAGAGTCAAAACCGCTTCCGGGATGC 1680  
1787 AATTTTGTGAAGATTTTGTCAATGTATGATTAAGAGTCAAAACCGCTTCCGGGATGC 1846  
1681 CGCTGCAAGACACAGTGAACACCAAGAGTCCCGTGTACCTGTGCTGAGAGTGT 1740  
1847 CGCTGCAAGACACAGTGAACACCAAGAGTCCCGTGTACCTGTGCTGAGAGTGT 1906  
1741 GACCCTGACCTCTGTCTTAATCTTGTGAGAGCGCTGACCAATGAGGACATGAATAATGTC 1800  
1907 GACCCTGACCTCTGTCTTAATCTTGTGAGAGCGCTGACCAATGAGGACATGAATAATGTC 1966  
1801 TGCAAGAACTGCAATTAACAGCGGAGCTCAAAAGACATTAATGCTGCAACATCTGAC 1860  
1967 TGCAAGAACTGCAATTAACAGCGGAGCTCAAAAGACATTAATGCTGCAACATCTGAC 2026  
1861 GTGCGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAGAAATTAATGATCTCAGAA 1920  
2027 GTGCGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAGAAATTAATGATCTCAGAA 2086  
1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAAAGTATGATTA 1980



841 TTTCATACGCTTTTCTGTAGGCGATGTTTTAATATAGCTGCTTCCATACCTTTTAT 900  
| | | | |  
Db 898 TTTCATACGCTTTTCTGTAGGCGATGTTTTAATATAGCTGCTTCCATACCTTTTAT 957  
| | | | |  
Qy 901 GCAACACCCCAACCTTATTAAGCGGAAGAACACAGAAACAGCTCTAGACAAACCTTGT 960  
| | | | |  
Db 958 GCAACACCCCAACCTTATTAAGCGGAAGAACACAGAAACAGCTCTAGACAAACCTTGT 1017  
| | | | |  
Qy 961 GGAACCAAGTGTACAGAGATTTGGAGGAGCAAGAGATTTGCTGCTGCTACCGCT 1020  
| | | | |  
Db 1018 GGAACCAAGTGTACAGAGATTTGGAGGAGCAAGAGATTTGCTGCTGCTACCGCT 1077  
| | | | |  
Qy 1021 GAGCGGATTAAGACCCCAACCAAAACCTCCAGAGGCGGAGAGAGAGAGGCTTCCCAT 1080  
| | | | |  
Db 1078 GAGCGGATTAAGACCCCAACCAAAACCTCCAGAGGCGGAGAGAGAGAGGCTTCCCAT 1137  
| | | | |  
Qy 1081 AACAGTAGAGGCGCCAGACCCCCACATTAAATGTGCTGGAATCAAGAGATACAGACGT 1140  
| | | | |  
Db 1138 AACAGTAGAGGCGCCAGACCCCCACATTAAATGTGCTGGAATCAAGAGATACAGACGT 1197  
| | | | |  
Qy 1141 GATAGGAGAGCGAGGAGCTGAAACGCGGGGAGAGAACATAGATTAAGAGAGAGAGAG 1200  
| | | | |  
Db 1198 GATAGGAGAGCGAGGAGCTGAAACGCGGGGAGAGAACATAGATTAAGAGAGAGAGAG 1257  
| | | | |  
Qy 1201 AAAGATGAACCTTCGAGCTCCTCTGAAGCAAAATTCCTGGGTCAAAACCAATTAAGATG 1260  
| | | | |  
Db 1258 AAAGATGAACCTTCGAGCTCCTCTGAAGCAAAATTCCTGGGTCAAAACCAATTAAGATG 1317  
| | | | |  
Qy 1261 AAGCCAAATATTGAACCTCCTGAGATGTGAGATGATGCTGAGACCTCAATGTTT 1320  
| | | | |  
Db 1318 AAGCCAAATATTGAACCTCCTGAGATGTGAGATGATGCTGAGACCTCAATGTTT 1377  
| | | | |  
Qy 1321 AAGATCTCTATGAGCACTTACTATAGACATTTCTGTGCAATTTGCTTAATTTGGAGAC 1380  
| | | | |  
Db 1378 AAGATCTCTATGAGCACTTACTATAGACATTTCTGTGCAATTTGCTTAATTTGGAGAC 1437  
| | | | |  
Qy 1381 AAAACATGTAGACAGGTATGATGTTTAAAGTCAAAAGATCTAGACATCACTACCTCAGCT 1440  
| | | | |  
Db 1438 AAAACATGTAGACAGGTATGATGTTTAAAGTCAAAAGATCTAGACATCACTACCTCAGCT 1497  
| | | | |  
Qy 1441 CCGGCTGAGAGATGTGATATCTCTCCAGAGAAAGAGAGAGAAACACCGGTTGTGGCT 1500  
| | | | |  
Db 1498 CCGGCTGAGAGATGTGATATCTCTCCAGAGAAAGAGAGAGAAACACCGGTTGTGGCT 1557  
| | | | |  
Qy 1501 GACACCTGCAAGAAATAGAGCTGAAGAAAGAGACGCTCTCTTAACATTTTCAACTAT 1560  
| | | | |  
Db 1558 GACACCTGCAAGAAATAGAGCTGAAGAAAGAGACGCTCTCTTAACATTTTCAACTAT 1617  
| | | | |  
Qy 1561 CAACCTGTGATCATCCACGGAGCCTGTGACAGTGTGCGCTTGTGTATAGACACA 1620  
| | | | |  
Db 1618 CAACCTGTGATCATCCACGGAGCCTGTGTGACAGTGTGCGCTTGTGTATAGACACA 1677  
| | | | |  
Qy 1621 AATTTTGTGAAGATTTTGTCAATGTATGATCAAGATGCAAAACCGCTTTCGGGATGC 1680  
| | | | |  
Db 1678 AATTTTGTGAAGATTTTGTCAATGTATGATCAAGATGCAAAACCGCTTTCGGGATGC 1737  
| | | | |  
Qy 1681 CGCTGCAAGACAGCTGCAACCAAGAGTCCCGTGTACTCTGCTGTCCGAGAGTGT 1740  
| | | | |  
Db 1738 CGCTGCAAGACAGCTGCAACCAAGAGTCCCGTGTACTCTGCTGTCCGAGAGTGT 1797  
| | | | |  
Qy 1741 GACCTGTGATCTGTCTTACTGTGTGAGGCGCTGACCAATGGGAGAGATTAATGTGTCC 1800  
| | | | |  
Db 1798 GACCTGTGATCTGTCTTACTGTGTGAGGCGCTGACCAATGGGAGAGATTAATGTGTCC 1857  
| | | | |  
Qy 1801 TGCAGAACTGCAATTCAGCGGAGCTCAAAAGAGATCTAATGCTGCGACCATCTGAC 1860  
| | | | |  
Db 1858 TGCAGAACTGCAATTCAGCGGAGCTCAAAAGAGATCTAATGCTGCGACCATCTGAC 1917  
| | | | |  
Qy 1861 GTGGAGAGCTGGGGATTTTATCAAAAGATCTGTGCGAGAAAGATTAATCTCTCAGAA 1920  
| | | | |  
Db 1918 GTGGAGAGCTGGGGATTTTATCAAAAGATCTGTGCGAGAAAGATTAATCTCTCAGAA 1977  
| | | | |  
Qy 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTGTATGATAA 1980  
| | | | |

Db 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTGTATGATAA 2037  
| | | | |  
Qy 1981 TACATGTGAGAGCTTTCTGTTCACCTTGAACATTAATTTGTGTGATGATCAACCGCAG 2040  
| | | | |  
Db 2038 TACATGTGAGAGCTTTCTGTTCACCTTGAACATTAATTTGTGTGATGATCAACCGCAG 2097  
| | | | |  
Qy 2041 GGTAAACAAATTCGTTTGCAAATCATTCGGTAATTCCAAACTGCTATGCAAAAGTTATG 2100  
| | | | |  
Db 2098 GGTAAACAAATTCGTTTGCAAATCATTCGGTAATTCCAAACTGCTATGCAAAAGTTATG 2157  
| | | | |  
Qy 2101 ATGTAAACGCTGATCAGAGATAGTATTTTGGCAAGAGAGCCATCCAGCTGGCGAA 2160  
| | | | |  
Db 2158 ATGTAAACGCTGATCAGAGATAGTATTTTGGCAAGAGAGCCATCCAGCTGGCGAA 2217  
| | | | |  
Qy 2161 GAGCTGTTTGTGATTAAGATTAACAGTACAGCAGGCTGATGCCCTGAATATGTGCGATCGAA 2220  
| | | | |  
Db 2218 GAGCTGTTTGTGATTAAGATTAACAGTACAGCAGGCTGATGCCCTGAATATGTGCGATCGAA 2277  
| | | | |  
Qy 2221 AGAGAAATGGAATTCCTTGA 2241  
| | | | |  
Db 2278 AGAGAAATGGAATTCCTTGA 2298  
| | | | |  
RESULT 9  
ABX76182  
ID ABX76182 standard; DNA; 2576 BP.  
XX  
XX ABX76182;  
XX  
DT 02-APR-2003 (first entry)  
XX  
XX Lung cancer-associated polynucleotide #24.  
DE  
XX Lung cancer-associated polynucleotide; gene; ds; cytosolic; emphysema;  
KW anti-inflammatory; antiaesthetic; non-small cell lung cancer; atelectasis;  
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX  
OS Unidentified.  
XX  
XX WO200286443-A2.  
XX  
XX 31-OCT-2002.  
PD  
XX 18-APR-2002; 2002WO-US012476.  
PF  
XX  
XX 18-APR-2001; 2001US-0284770P.  
PR 10-MAY-2001; 2001US-0280492P.  
PR 09-NOV-2001; 2001US-0339245P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 29-NOV-2001; 2001US-0334370P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
XX (BOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Aziz N. Murray R;  
PI  
XX  
XX MPI; 2003-093161/08.  
DR P-PSDB; ABUS6461.  
XX  
XX  
XX Detecting a lung cancer-associated transcript in a cell from a patient  
PT for treating lung cancer, by contacting a biological sample from the  
PT patient with a polynucleotide that exhibits increased or decreased  
PT expression in lung cancer.  
XX  
XX Claim 22; Page 231-232; 453pp; English.  
XX  
XX The invention relates to a method for detecting a lung cancer-associated  
CC transcript in a cell from a patient, comprising contacting a biological  
CC sample from the patient with a polynucleotide that selectively hybridises  
CC to a sequence that is at least 80 % identical to a gene that exhibits

increased or decreased expression in lung cancer samples. Lung cancer-associated polynucleotides and polypeptides are used for identifying a compound that modulates a lung cancer-associated polypeptide, for inhibiting proliferation of a lung cancer-associated cell to treat lung cancer in a patient and for treating a mammal having lung cancer by administering a modulatory compound identified. The methods are useful for treating lung cancer, such as small cell lung cancer, non-small cell lung cancer or other benign or precancerous lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis, hyperresponsivity pneumonitis, interstitial pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides and polypeptides are useful for diagnostic purposes and as targets for screening for therapeutic compounds that modulate lung cancer, such as antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the invention

Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Query Match 99.8%; Score 2236.2; DB 8; Length 2576;

Best Local Similarity 99.9%; Pred. No. 0; Mismatches 3; Indels 0; Gaps 0;

Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 ATGGGCGCATCTGGGAAAGAAATCTGAAGAGGACCAAGTTTGTGGCGGAACGCTGTAAA 60
DB 58 ATGGGCGCATCTGGGAAAGAAATCTGAAGAGGACCAAGTTTGTGGCGGAACGCTGTAAA 117
OY 61 TCAGAGTACATGCGCATGAGACAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGAGT 120
DB 118 TCAGAGTACATGCGCATGAGACAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGAGT 177
OY 121 ATGTTAGTCCAAATCGTCAGAAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180
DB 178 ATGTTAGTCCAAATCGTCAGAAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 237
OY 181 AAACGCGCAAGATACAGCTGTGTGACATCTGTAATCTTGTGAAGTCAATTCGGCGGACT 240
DB 238 AAACGCGCAAGATACAGCTGTGTGACATCTGTAATCTTGTGAAGTCAATTCGGCGGACT 297
OY 241 AGGAGTGTTCGGTACCAAGTACATGTTTCCAAACAAGTATCCCATTTAAAGACT 300
DB 298 AGGAGTGTTCGGTACCAAGTACATGTTTCCAAACAAGTATCCCATTTAAAGACT 357
OY 301 CTGAATGCAATGCTCTCAGTACCAATATGTAATTTGGTCTCCCTTACAGAGAAATTT 360
DB 358 CTGAATGCAATGCTCTCAGTACCAATATGTAATTTGGTCTCCCTTACAGAGAAATTT 417
OY 361 ATGTGTGAAGATGAACCTGTTTACATAACATTCCTTAAATGGGAGATGAAGTTTAAAT 420
DB 418 ATGTGTGAAGATGAACCTGTTTACATAACATTCCTTAAATGGGAGATGAAGTTTAAAT 477
OY 421 CAGAGTGTATCTTCATTTGAAGAACTAATTAATTAATGATGGAAAGTACACGGAGAT 480
DB 478 CAGAGTGTATCTTCATTTGAAGAACTAATTAATTAATGATGGAAAGTACACGGAGAT 537
OY 481 AGAGATGTGGGTATTATAATGATGAATAATTTTGTGAAGTGTGAATGCTTGTGTCAA 540
DB 538 AGAGATGTGGGTATTATAATGATGAATAATTTTGTGAAGTGTGAATGCTTGTGTCAA 597
OY 541 TATATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 598 TATATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
OY 601 AAAAGTGTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTTCTTCT 660
DB 658 AAAAGTGTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTTCTTCT 717
OY 661 GATTAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGAGCAAGCAGAGAACTA 720
DB 718 GATTAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGAGCAAGCAGAGAACTA 777
OY 721 AAGGAAATATTAAGAACTCACCGAAGCAGCTTCCAGGCGGCACTTCTCTGTAATGT 780
DB 778 AAGGAAATATTAAGAACTCACCGAAGCAGCTTCCAGGCGGCACTTCTCTGTAATGT 837
```

```
OY 781 ACCCCCAACATAGATGACCAAAATGCTAATCTGTTCAAGAGAGCAAAAGCTTACACTCC 840
DB 838 ACCCCCAACATAGATGACCAAAATGCTAATCTGTTCAAGAGAGCAAAAGCTTACACTCC 897
OY 841 TTTTCATAGCTTTTCTGTAGCCATGTTTAAATATGATCGCTTCCATACCTCTTTCAT 900
DB 898 TTTTCATAGCTTTTCTGTAGCCATGTTTAAATATGATCGCTTCCATACCTCTTTCAT 957
OY 901 GCACACCCCAACTTTAAGCGGAGAGACAGAAACAGCTCTAGACAAACCTTGT 960
DB 958 GCACACCCCAACTTTAAGCGGAGAGACAGAAACAGCTCTAGACAAACCTTGT 1017
OY 961 GGACCAAGTGTATCCAGCATTTTGAAGAGCAAAAGAGTTTGTCTGCTCTCACCGCT 1020
DB 1018 GGACCAAGTGTATCCAGCATTTTGAAGAGAGCAAAAGAGTTTGTCTGCTCTCACCGCT 1077
OY 1021 GAGCGATTAAGACCCCAACCAAAACGTTCCAGAGGCGCGAGAAAGAGAGCGGCTTCCCAAT 1080
DB 1078 GAGCGATTAAGACCCCAACCAAAACGTTCCAGAGGCGCGAGAAAGAGAGCGGCTTCCCAAT 1137
OY 1081 AACAGTAGAGGCGCCAGACACCCCAACCAATTAATGTGCTGGAATCAAGATACAGACGT 1140
DB 1138 AACAGTAGAGGCGCCAGACACCCCAACCAATTAATGTGCTGGAATCAAGATACAGACGT 1197
OY 1141 GATAGGAGAGAGGAGCTGAAAACGGGGGAGAGAACATATTAAGAAAGAAAGAGAG 1200
DB 1198 GATAGGAGAGAGGAGCTGAAAACGGGGGAGAGAACATATTAAGAAAGAAAGAGAG 1257
OY 1201 AAAGATGAATCTTGACCTCTCTGAGCAAAATCTCGGTGTCAAACCAATTAAGATG 1260
DB 1258 AAAGATGAATCTTGACCTCTCTGAGCAAAATCTCGGTGTCAAACCAATTAAGATG 1317
OY 1261 AAGCCAAATATTGAACCTCTGAGAAATGTGAGTGTGAGTGTGAGAGCTCAATGTTT 1320
DB 1318 AAGCCAAATATTGAACCTCTCTGAGAAATGTGAGTGTGAGTGTGAGAGCTCAATGTTT 1377
OY 1321 AGAGTCTCATTTGCACTTATGATGAATTTCTGTGCCATTTGCTTAAATTTGGAGAC 1380
DB 1378 AGAGTCTCATTTGCACTTATGATGAATTTCTGTGCCATTTGCTTAAATTTGGAGAC 1437
OY 1381 AAAACATGTGACAGGTGTATGAGTTTAAAGTCAAAATCTAGCATCATAGCTCCAGCT 1440
DB 1438 AAAACATGTGACAGGTGTATGAGTTTAAAGTCAAAATCTAGCATCATAGCTCCAGCT 1497
OY 1441 CCCGCTGAGATGTGATTAATCTCTCAAGAAAAAGAGAAACAACCGGTTGTGGCT 1500
DB 1498 CCCGCTGAGATGTGATTAATCTCTCAAGAAAAAGAGAAACAACCGGTTGTGGCT 1557
OY 1501 GCAACATGCAAGAAAGATACAGCTGAAAAGAGACGGCTCTTAAACATGTTTACATAT 1560
DB 1558 GCAACATGCAAGAAAGATACAGCTGAAAAGAGACGGCTCTTAAACATGTTTACATAT 1617
OY 1561 CAACCTGTGATCAATCCACGGCAACCTTGTGACAGTTTGGCCCTGTGTGATGACAA 1620
DB 1618 CAACCTGTGATCAATCCACGGCAACCTTGTGACAGTTTGGCCCTGTGTGATGACAA 1677
OY 1621 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAAACCGCTTTCGGGATGC 1680
DB 1678 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAAACCGCTTTCGGGATGC 1737
OY 1681 CGCTGCAAAAGACAGTGTCAACCAAGAGAGTCCCGGTCTACCTGCTGTCCGAGAGTGT 1740
DB 1738 CGCTGCAAAAGACAGTGTCAACCAAGAGAGTCCCGGTCTACCTGCTGTCCGAGAGTGT 1797
OY 1741 GACCTTGAACCTCTGCTTAACTTGTGAGCGGCTGACCAATGGGACAGTAAATATGTGCC 1800
DB 1798 GACCTTGAACCTCTGCTTAACTTGTGAGCGGCTGACCAATGGGACAGTAAATATGTGCC 1857
OY 1801 TGCAAGAACTGCAATATTCAGCGGGGCTCAAAAAGCATTTATGCTGTGACCAATCTGAC 1860
DB 1858 TGCAAGAACTGCAATATTCAGCGGGGCTCAAAAAGCATTTATGCTGTGACCAATCTGAC 1917
```

QY 1861 GTGGAGGCTGGGGATTTTATTCAGAAAGTCGTGACAGAAAAATGATTCATCTCAGA 1920  
DB 1918 GTGGAGGCTGGGGATTTTATTCAGAAAGTCGTGACAGAAAAATGATTCATCTCAGA 1977  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGAGAAAGTATGATTA 1980  
DB 1978 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGAGAAAGTATGATTA 2037  
QY 1981 TACATGTGACGCTTTCTGTTCACCTTGAACATGATTTGTGTGATGCAACCCGCAAG 2040  
DB 2038 TACATGTGACGCTTTCTGTTCACCTTGAACATGATTTGTGTGATGCAACCCGCAAG 2097  
QY 2041 GGTACCAAAATCGTTTTCATCATTCGTGTAATTCAGAACTGATGCAAAAGTTATG 2100  
DB 2098 GGTACCAAAATCGTTTTCATCATTCGTGTAATTCAGAACTGATGCAAAAGTTATG 2157  
QY 2101 ATGTTAAACGTGATCAGATAGATATTTTTCAGAGAGCCATCCAGATCGAGCA 2160  
DB 2158 ATGTTAAACGTGATCAGATAGATATTTTTCAGAGAGCCATCCAGATCGAGCA 2217  
QY 2161 GAGCTGTTTTCATTCAGATACAGCCAGCTGATGCTGATGATGTCGATTCGAA 2220  
DB 2218 GAGCTGTTTTCATTCAGATACAGCCAGCTGATGCTGATGATGTCGATTCGAA 2277  
QY 2221 AGAGAAATGGAATCCCTTGA 2241  
DB 2278 AGAGAAATGGAATCCCTTGA 2298

RESULT 10  
ADK66988  
ID ADK66988 standard; DNA; 2576 BP.  
XX  
AC ADK66988;  
XX

DT 06-MAY-2004 (first entry)  
XX

DE Gene #8 for inhibitory RNA to manipulate stem cell phenotype.  
XX

KW ds; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
XX  
KW pluripotent stem cell.  
XX

OS Homo sapiens.  
XX

PN WO2003068961-A2.  
XX

PD 21-AUG-2003.  
XX

PE 12-FEB-2003; 2003WO-GB000579.  
XX

PR 13-FEB-2002; 2002GB-00003359.  
XX

PR 13-FEB-2002; 2002GB-00003387.  
XX

PA (AXOR-) AXORDIA LTD.  
XX

PI Andrews P, Walsh J, Gokhale P;  
XX

DR WPI; 2003-697528/66.  
XX

PT New inhibitory RNA molecule having double stranded RNA molecules, useful  
XX  
PT for manipulating the phenotype of stem cells, preferably pluripotent  
XX  
PT stem cells.  
XX

PS Disclosure; SEQ ID NO 78; 157bp; English.  
XX

CC The invention relates to an inhibitory RNA (RNAi) molecule derived from a  
XX  
CC nucleic acid molecule comprising a defined nucleic acid sequences given  
XX  
CC in the specification or a sequence which hybridizes to the sequences and  
XX  
CC encodes a Notch signaling target gene or which is a degenerate as a  
XX  
CC result of the genetic code of the sequences. The methods and compositions  
XX  
CC of the present invention are useful for manipulating the phenotype of  
XX  
CC stem cells, preferably pluripotent stem cells. This sequence corresponds  
XX  
CC to one of the nucleic acid molecules of the invention.

XX  
SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Query Match 99.8%; Score 2236.2; DB 10; Length 2576;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGCCAGACTGGGAAGAAATCTGAGAAAGGACACAGTTGTTGGCGGAAGCGTGTAA 60  
DB 58 ATGGGCCAGACTGGGAAGAAATCTGAGAAAGGACACAGTTGTTGGCGGAAGCGTGTAA 117  
QY 61 TCAGAGTACATGCCGACTGACAGAGCTCAAGAGTTTCAGACGAGCTGATGAAAGTAAAGT 120  
DB 118 TCAGAGTACATGCCGACTGACAGAGCTCAAGAGTTTCAGACGAGCTGATGAAAGTAAAGT 177  
QY 121 ATGTTAGTTCATCGTACAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180  
DB 178 ATGTTAGTTCATCGTACAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 237  
QY 181 AAACGCGAAGATACAGCCCTGTGACATCTGAGCTTCTGTGAGCTCATTTGGCGGAGCT 240  
DB 238 AAACGCGAAGATACAGCCCTGTGACATCTGAGCTTCTGTGAGCTCATTTGGCGGAGCT 297  
QY 241 AGGAGTGTTCGCTGACAGTGAATTTTCCAAACAAGTATCCCATTTAAAGACT 300  
DB 298 AGGAGTGTTCGCTGACAGTGAATTTTCCAAACAAGTATCCCATTTAAAGACT 357  
QY 301 CTGAATGAGTTCCTCAGTACCAATTAATGATTTCTGCTCCCTTACAGCAATTTT 360  
DB 358 CTGAATGAGTTCCTCAGTACCAATTAATGATTTCTGCTCCCTTACAGCAATTTT 417  
QY 361 ATGTTGGAAGATGAAACGTTTTCATTAACATCTCTTATATGGAGATGAAGTTTAAAT 420  
DB 418 ATGTTGGAAGATGAAACGTTTTCATTAACATCTCTTATATGGAGATGAAGTTTAAAT 477  
QY 421 CAGATGCTACTTCATTCATGAAAGAACTAATTAATAATTTATGATGGGAAAGTACACGGAGT 480  
DB 478 CAGATGCTACTTCATTCATGAAAGAACTAATTAATAATTTATGATGGGAAAGTACACGGAGT 537  
QY 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGATATGCTTGGTCA 540  
DB 538 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGATATGCTTGGTCA 597  
QY 541 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 598 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 601 AAAGATCGAGGATCAACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCT 660  
DB 658 AAAGATCGAGGATCAACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCT 717  
QY 661 GATAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGACTA 720  
DB 718 GATAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGACTA 777  
QY 721 AAAGAAATATTAAGAACTCAACCGAGATGATTAAGAAAGCCGCCACCTCTCTGAATGT 780  
DB 778 AAAGAAATATTAAGAACTCAACCGAGATGATTAAGAAAGCCGCCACCTCTCTGAATGT 837  
QY 781 ACCCCCAACATGATGACCAATGCTTAATCTGTTGAGAGAGCAAAAGCTTACACCTC 840  
DB 838 ACCCCCAACATGATGACCAATGCTTAATCTGTTGAGAGAGCAAAAGCTTACACCTC 897  
QY 841 TTTCATACGCTTTTCTGAGCGATGTTTAAATATGATCTGCTTCTCATCTTTTCAT 900  
DB 898 TTTCATACGCTTTTCTGAGCGATGTTTAAATATGATCTGCTTCTCATCTTTTCAT 957  
QY 901 GCAACACCCCAACATTTATAGCGGAGAAACACAGAAACGCTCTTACACAAACCTTGT 960  
DB 958 GCAACACCCCAACATTTATAGCGGAGAAACACAGAAACGCTCTTACACAAACCTTGT 1017  
QY 961 GGACCAACGTTTACGACGATTTTGAGGGAGCAAGAGTTTGCTGCTCAGCAGCT 1020









Db 1858 TCGAAGAACTGACAGTATTACAGCGGGCTCCAAAAGCATCTATTGCTGCGACCATCTGAC 1917  
 Qy 1861 GTGGCAGGCTGGGGATTTTATCATAGATCTGTGCAGAAAAATGAAATTCATCTCAGAA 1920  
 Db 1918 GTGGCAGGCTGGGGATTTTATCATAGATCTGTGCAGAAAAATGAAATTCATCTCAGAA 1977  
 Qy 1921 TACTGTGAGAGATTTATTTCTCAGATGAGCTGCAGAGAGGAAAGTGTATGATPAA 1980  
 Db 1978 TACTGTGAGAGATTTATTTCTCAGATGAGCTGCAGAGAGGAAAGTGTATGATPAA 2037  
 Qy 1981 TACATGTGACGCTTCTGTCAACTGTAACAATGATTTGTGTGATGCAACCGCAG 2040  
 Db 2038 TACATGTGACGCTTCTGTCAACTGTAACAATGATTTGTGTGATGCAACCGCAG 2097  
 Qy 2041 GGTAAACAATTCGTTTGGCAATCATTCGGTAATCCAACTGCTATGCAAAAGTTATG 2100  
 Db 2098 GGTAAACAATTCGTTTGGCAATCATTCGGTAATCCAACTGCTATGCAAAAGTTATG 2157  
 Qy 2101 ATGTTAAGCGTATCAGAGATGATATTTTGGCAAGAGCCATCAGACTGCGGAA 2160  
 Db 2158 ATGTTAAGCGTATCAGAGATGATATTTTGGCAAGAGCCATCAGACTGCGGAA 2217  
 Qy 2161 GAGCTGTTTGTATTAAGATACAGCAGGCTGATGCTGAAAGTATGCGGATGGA 2220  
 Db 2218 GAGCTGTTTGTATTAAGATACAGCAGGCTGATGCTGAAAGTATGCGGATGGA 2277  
 Qy 2221 AGAGAAATGGAATCCCTTGA 2241  
 Db 2278 AGAGAAATGGAATCCCTTGA 2298

RESULT 12  
 ID ADR25182 standard; DNA; 2576 BP.  
 AC ADR25182;  
 DT 21-OCT-2004 (first entry)  
 DE Breast cancer prognosis marker #1043.  
 KM ds; breast cancer; prognosis; gene expression; diagnosis.  
 OS Homo sapiens.  
 PN WO2004065545-A2.  
 PD 05-AUG-2004.  
 PF 15-JAN-2004; 2004WO-US001100.  
 PR 15-JAN-2003; 2003US-00342887.  
 PA (ROSE-) ROSETTA INPHARMATICS LLC.  
 PI (NECA-) NETHERLANDS CANCER INST.  
 PI Van't Veer LJ, He Y;  
 DR WPI; 2004-593473/57.  
 PT Classifying a breast cancer patient according to prognosis comprises  
 PT determining the similarity between the level of expression of each of  
 PT five genes in a cell sample taken from patient, to control levels.  
 XX Disclosure; SEQ ID NO 1043; 226bp; English.  
 CC The invention relates to a method of classifying a breast cancer patient  
 CC according to prognosis by determining the similarity between the level of  
 CC expression of each of five genes for which markers are listed in the  
 CC specification, in a cell sample taken from the breast cancer patient, to  
 CC control levels of expression for each respective five genes to obtain a  
 CC patient similarity value. The methods are useful for classifying a breast  
 CC cancer patient according to prognosis. Kits and computer program products

CC are useful for data analysis using the diagnostic, prognostic and  
 CC statistical methods of the invention. This sequence corresponds to a  
 CC marker used in the method of the invention.  
 XX  
 SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 2236.2; DB 13; Length 2576;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 ATGGGCGCAGCTGGGAGAAATCTGAGAGGAGCAAGTTTGTGGCGGAAAGCTGTAAA 60  
 Db 58 ATGGGCGCAGCTGGGAGAAATCTGAGAGGAGCAAGTTTGTGGCGGAAAGCTGTAAA 117  
 Qy 61 TCAAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGAGCTGATGAATGAAGT 120  
 Db 118 TCAAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGAGCTGATGAATGAAGT 177  
 Qy 121 ATGTTTGTCCCAATCGTCAGAAAATTTTGGAAAAGAGGAAATCTTAAACCAAGATGG 180  
 Db 178 ATGTTTGTCCCAATCGTCAGAAAATTTTGGAAAAGAGGAAATCTTAAACCAAGATGG 237  
 Qy 181 AAACAGGAGAGATACAGCTGTGCAATCTGAGCTTCTGTGAGCTCATTTGCGGGACT 240  
 Db 238 AAACAGGAGAGATACAGCTGTGCAATCTGAGCTTCTGTGAGCTCATTTGCGGGACT 297  
 Qy 241 AGGAGTGTGCGTACAGCTGATGATTTTCCAAACAAGTATCCATTTAAAGACT 300  
 Db 298 AGGAGTGTGCGTACAGCTGATGATTTTCCAAACAAGTATCCATTTAAAGACT 357  
 Qy 301 CTGAATGAGTGTGCTGATGATCCATATGATCTTCTGCTCCCTACAGCAGAAATTTT 360  
 Db 358 CTGAATGAGTGTGCTGATGATCCATATGATCTTCTGCTCCCTACAGCAGAAATTTT 417  
 Qy 361 ATGTTGAGAGATGAACTGTTTACATTAATCTTATATGAGAGATGAAGTTTATGAT 420  
 Db 418 ATGTTGAGAGATGAACTGTTTACATTAATCTTATATGAGAGATGAAGTTTATGAT 477  
 Qy 421 CAGATGATGATCTTCAATTAAGAACTTATTAAGAAATTTATGATGGAAGATACAGGGAT 480  
 Db 478 CAGATGATGATCTTCAATTAAGAACTTATTAAGAAATTTATGATGGAAGATACAGGGAT 537  
 Qy 481 AGAAGATGAGGTTTATTAATGATGAATTTTGTGAGTTGTGGAATGCGCTTGTCAA 540  
 Db 538 AGAAGATGAGGTTTATTAATGATGAATTTTGTGAGTTGTGGAATGCGCTTGTCAA 597  
 Qy 541 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 Db 598 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 Qy 601 AAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 Db 658 AAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
 Qy 661 GATAAAATTTTGAAGCATTTTCTCAATGTTTCAATTAAGGCAAGCAAGCAAGCAACTA 720  
 Db 718 GATAAAATTTTGAAGCATTTTCTCAATGTTTCAATTAAGGCAAGCAAGCAAGCAACTA 777  
 Qy 721 AAGGAAAAATTAAGAACTCAAGCAAGCAAGCTCCAGCGCATCTTCTCTGAATGT 780  
 Db 778 AAGGAAAAATTAAGAACTCAAGCAAGCAAGCTCCAGCGCATCTTCTCTGAATGT 837  
 Qy 781 ACCCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 Db 838 ACCCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
 Qy 841 TTTGATTCGCTTCTGATGCGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 900  
 Db 898 TTTGATTCGCTTCTGATGCGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 957  
 Qy 901 GCAACACCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 Db 958 GCAACACCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1017



CC sample it is correlated with a poor prognosis for breast cancer, given in  
CC an exemplification of the invention.

XX Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Query Match 99.8%; Score 2236.2; DB 14; Length 2576;

Best Local Similarity 99.9%; Fred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATGGGCCGAGCTGGGAAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAACGCTGTAAA 60
DB 58 ATGGGCCGAGCTGGGAAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAACGCTGTAAA 117
QY 61 TCAGAGTACATCGCATGTGAGCAAGCTCAAGAGTTGAGCGAGTGAAGTAAGT 120
DB 118 TCAGAGTACATCGCATGTGAGCAAGCTCAAGAGTTGAGCGAGTGAAGTAAGT 177
QY 121 ATGTTTAACTTCATTCGTCAAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATG 180
DB 178 ATGTTTAACTTCATTCGTCAAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATG 237
QY 181 AAACGCGAAGATATACAGCTGTGACATCTGACATCTGTGAGCTCATTTGCGGAGCT 240
DB 238 AAACGCGAAGATATACAGCTGTGACATCTGACATCTGTGAGCTCATTTGCGGAGCT 297
QY 241 AGGAGTGTTCGGTGAACCAAGTGAATTTTCCAAACAAGTCAATCCATTAAAGACT 300
DB 298 AGGAGTGTTCGGTGAACCAAGTGAATTTTCCAAACAAGTCAATCCATTAAAGACT 357
QY 301 CTGAATGAGTTGCTTCAGTACCCCAATATGTTCTGTCCTCCCTACAGAGAAATTT 360
DB 358 CTGAATGAGTTGCTTCAGTACCCCAATATGTTCTGTCCTCCCTACAGAGAAATTT 417
QY 361 ATGTGGAAGATGAAACTGTTTACATTAACATTCCTTATATGAGAGATGAAGTTTATAGT 420
DB 418 ATGTGGAAGATGAAACTGTTTACATTAACATTCCTTATATGAGAGATGAAGTTTATAGT 477
QY 421 CAGATGTGACTTTCATTGAAGAACTAATAAATAATATGATGGAAAGTACACGGAGAT 480
DB 478 CAGATGTGACTTTCATTGAAGAACTAATAAATAATATGATGGAAAGTACACGGAGAT 537
QY 481 AGAGAAATGGGTTTATTAAGTAAATTTTGTGGAGTTGGTGAATGCCCTTGCTGTA 540
DB 538 AGAGAAATGGGTTTATTAAGTAAATTTTGTGGAGTTGGTGAATGCCCTTGCTGTA 597
QY 541 TATAATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 598 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
QY 601 AAAGATCTGAGAGATCACCGAGATGATAAGAAAGCCGCCCACTCGGAAATTTCTTTCT 660
DB 658 AAAGATCTGAGAGATCACCGAGATGATAAGAAAGCCGCCCACTCGGAAATTTCTTTCT 717
QY 661 GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACACGAGAAAGACTA 720
DB 718 GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACACGAGAAAGACTA 777
QY 721 AAGGAAAAATTAAGAACTCACCGAAGAGAGCTCCGAGGCGCACTTCTCTGTAATGT 780
DB 778 AAGGAAAAATTAAGAACTCACCGAAGAGAGCTCCGAGGCGCACTTCTCTGTAATGT 837
QY 781 ACCCCCAACATAGATGACCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840
DB 838 ACCCCCAACATAGATGACCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 897
QY 841 TTTGATAGCTTTTCTGTAGGAGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 900
DB 898 TTTGATAGCTTTTCTGTAGGAGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 957
QY 901 GCAACACCAACACTTATTAAGCGAAGAACACAGAAAGAGCTTGAACAACAACTTGT 960
DB 958 GCAACACCAACACTTATTAAGCGAAGAACACAGAAAGAGCTTGAACAACAACTTGT 1017
```

```

QY 961 GACACCAAGTTTACACGACATTTTGAAGGAGCAAAAGAGTTTGTGCTGCTTCAACGGCT 1020
DB 1018 GACACCAAGTTTACACGACATTTTGAAGGAGCAAAAGAGTTTGTGCTGCTTCAACGGCT 1077
QY 1021 GAGCGATTAAGACCCCAACAAAACGTCACAGAGGCGCGACAGAAAGAGAGCGCTTCCCAAT 1080
DB 1078 GAGCGATTAAGACCCCAACAAAACGTCACAGAGGCGCGACAGAAAGAGAGCGCTTCCCAAT 1137
QY 1081 AACAGTAGACGCGCCAGACACCCCAACCAATTAATGTGCTGAAATCAAAAGATATACAGACT 1140
DB 1138 AACAGTAGACGCGCCAGACACCCCAACCAATTAATGTGCTGAAATCAAAAGATATACAGACT 1197
QY 1141 GATAGGAAACAGGAGATTTGAAAACGGGGGAGAGAACAAATATAAGAAAGAAAGAGAG 1200
DB 1198 GATAGGAAACAGGAGATTTGAAAACGGGGGAGAGAACAAATATAAGAAAGAAAGAGAG 1257
QY 1201 AAAGATGAAACTTGAAGCTCCTCTGAAGCAAAATTTCTGCTGTCAAAACCAATTAAGATG 1260
DB 1258 AAAGATGAAACTTGAAGCTCCTCTGAAGCAAAATTTCTGCTGTCAAAACCAATTAAGATG 1317
QY 1261 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGAGTGTGAAGCTTCAATGTTT 1320
DB 1318 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGATGTGAGTGTGAAGCTTCAATGTTT 1377
QY 1321 AGAGTCTCTATTTGCACTTATGATGAATTTTCTGTGSCCAATTTGCTAGTTTAAATTGGAGC 1380
DB 1378 AGAGTCTCTATTTGCACTTATGATGAATTTTCTGTGSCCAATTTGCTAGTTTAAATTGGAGC 1437
QY 1381 AAAACATGTAGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1438 AAAACATGTAGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
QY 1441 CCGGCTGAGAGTGTGATCTCTCCAAAGAAAAGAAAGAAAACACCGGTTGTGGGCT 1500
DB 1498 CCGGCTGAGAGTGTGATCTCTCCAAAGAAAAGAAAGAAAACACCGGTTGTGGGCT 1557
QY 1501 GCACATGTGCAAGAAAGATACAGCTGTAABAAAGCGGCTCCTAACCAATGTTTCAACTAT 1560
DB 1558 GCACATGTGCAAGAAAGATACAGCTGTAABAAAGCGGCTCCTAACCAATGTTTCAACTAT 1617
QY 1561 CAACCTGTGATCATCCACAGGACGCTTGTGACAGTTGCTGACCTTGTGTGATAGACAA 1620
DB 1618 CAACCTGTGATCATCCACAGGACGCTTGTGACAGTTGCTGACCTTGTGTGATAGACAA 1677
QY 1621 AATTTTGTGAAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1678 AATTTTGTGAAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATGAT 1737
QY 1681 CGCTGCAAGACAGTGCACACCAAGCAGGCGCGGCTACCTGGGCTGTCCGAGAGGT 1740
DB 1738 CGCTGCAAGACAGTGCACACCAAGCAGGCGCGGCTACCTGGGCTGTCCGAGAGGT 1797
QY 1741 GACCTGACCTCTGTCTTACTTGTGTGAGCGGCTGACCAATTTGGGACAGTAAATGTGTCC 1800
DB 1798 GACCTGACCTCTGTCTTACTTGTGTGAGCGGCTGACCAATTTGGGACAGTAAATGTGTCC 1857
QY 1801 TGCAAGAACTGCAATTAACAGCGGCGCTCCAAAAGCAATTAATGCTGTGAC 1860
DB 1858 TGCAAGAACTGCAATTAACAGCGGCGCTCCAAAAGCAATTAATGCTGTGAC 1917
QY 1861 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAAATGAATCATCTGAGAA 1920
DB 1918 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAAATGAATCATCTGAGAA 1977
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGATGATAA 1980
DB 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGATGATAA 2037
QY 1981 TACATGTGACAGCTTGTGTTCAACTGAACAATGATTTGTGTGATGACACCGCAAG 2040
DB 2038 TACATGTGACAGCTTGTGTTCAACTGAACAATGATTTGTGTGATGACACCGCAAG 2097
QY 2041 GGTAAACAAATTCGTTTGGCAATCATTCGGTAAATCCAAACCTGTAATGCAAAAAGTTATG 2100
```

Db 2098 GGTAAACAAATTCGTTTGCAATCATTCGTAATCCAACTGCTATGCAAAAGTTAAG 2157  
Qy 2101 ATGTTAAACGGATGACAGATAGGTATTTTGGCAAGAGCCATCCAGCTGCGGAA 2160  
Db 2158 ATGTTAAACGGATGACAGATAGGTATTTTGGCAAGAGCCATCCAGCTGCGGAA 2217  
Qy 2161 GAGCTGTTTGTGATTAAGATACAGCCGAGTGTGCTGATGCTTGAAGTATGCTGGCATTCGA 2220  
Db 2218 GAGCTGTTTGTGATTAAGATACAGCCGAGTGTGCTGATGCTTGAAGTATGCTGGCATTCGA 2277  
Qy 2221 AGAGAAATGGAATCCCTTGA 2241  
Db 2278 AGAGAAATGGAATCCCTTGA 2298  
RESULT 14  
ADD18523  
ID ADD18523 standard; DNA; 2512 BP.  
XX  
AC ADD18523;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human prostate cancer diagnosis related DNA sequence SegID95.  
XX  
KM prostate tissue; cancer diagnostic; cancer marker; prostate cancer; PCA;  
KM male cancer-related death; serum biomarker; tissue biomarker; cytostatic;  
KM gene therapy; prostate biopsy tissue; AMACR;  
KM alpha-methylacyl-coenzyme A racemase; diagnosing cancer; cell growth;  
KM human; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2003012067-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 02-AUG-2002; 2002WO-US024567.  
XX  
PR 02-AUG-2001; 2001US-0309581P.  
PR 15-NOV-2001; 2001US-0334468P.  
PR 01-AUG-2002; 2002US-00210120.  
XX  
PA (UNMT ) UNIV MICHIGAN.  
XX  
PI Rubin MA, Chinaiyan AM, Sreekumar A;  
XX  
DR WPI; 2003-278396/27.  
XX  
PT Characterizing prostate tissue comprises providing a prostate tissue  
PT sample from a subject and detecting the presence or absence of expression  
PT of hepsin, p1m-1 or EHz2.  
XX  
PS Disclosure; SEQ ID NO 95; 297bp; English.  
XX  
CC This invention relates to a novel method of characterizing prostate  
CC tissue in a subject and to compositions and methods for cancer  
CC diagnostics, including cancer markers, in particular prostate cancer.  
CC Prostate cancer (PCA) is a leading cause of male cancer-related death.  
CC Additional serum and tissue biomarkers would aid diagnosis. The invention  
CC may provide means of producing compounds with a cytostatic activity or  
CC allow the development of gene therapy. The methods of the invention  
CC useful for characterizing prostate tissue in a subject, screening  
CC compounds, characterizing inconclusive prostate biopsy tissue in a  
CC subject, detecting AMACR (alpha-methylacyl-coenzyme A racemase)  
CC expression in a bodily fluid, characterizing tissue in a subject,  
CC diagnosing cancer in a subject and inhibiting the growth of cells. The  
CC present sequence is a DNA sequence which is preferably utilised in the  
CC method of the invention.  
XX  
SQ Sequence 2512 BP; 798 A; 501 C; 588 G; 625 T; 0 U; 0 Other;

Query Match 99.5%; Score 2229; DB 10; Length 2512;  
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;  
Matches 2240; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
Qy 1 ATGGGCCAGACTGGGAGAGAAATCTGAGAGGAGCAAGTTTGTGGCGGAGGTAAAA 60  
Db 10 ATGGGCCAGACTGGGAGAGAAATCTGAGAGGAGCAAGTTTGTGGCGGAGGTAAAA 69  
Qy 61 TCAGAGTACATGCGCATGAGACAGCTCAAGAGTTTCAAGCAAGCTGATGAAGTAAAGT 120  
Db 70 TCAGAGTACATGCGCATGAGACAGCTCAAGAGTTTCAAGCAAGCTGATGAAGTAAAGT 128  
Qy 121 ATGTTAGTCCAAATGTCAGAAAATTTGGAAAAGAGCAAAATCTTAAACCAAGATAG 180  
Db 129 ATGTTAGTCCAAATGTCAGAAAATTTGGAAAAGAGCAAAATCTTAAACCAAGATAG 188  
Qy 181 AAACAGCAGAGATACAGCTGTGACATCTCTGACTTCTGTAGCTCATTTGCGCGGACT 240  
Db 189 AAACAGCAGAGATACAGCTGTGACATCTCTGACTTCTGTAGCTCATTTGCGCGGACT 248  
Qy 241 AGGAGTGTGCGTGAACAGTGAAGTGTGATTTTCCACACAAAGTATCCATTAAAGCT 300  
Db 249 AGGAGTGTGCGTGAACAGTGAAGTGTGATTTTCCACACAAAGTATCCATTAAAGCT 308  
Qy 301 CTGAATGCAATGCTTCACTGATCCATTAATGATCTGCTCCCTTCAAGAGAAATTT 360  
Db 309 CTGAATGCAATGCTTCACTGATCCATTAATGATCTGCTCCCTTCAAGAGAAATTT 368  
Qy 361 ATGGTGAAGATGAAGAACTGTTTTCATTAACATTCCTTATATGAGAGATGAAGTTTAGAT 420  
Db 369 ATGGTGAAGATGAAGAACTGTTTTCATTAACATTCCTTATATGAGAGATGAAGTTTAGAT 428  
Qy 421 CAGATGTACTTTTCATTTGAAGAACTAATTAATAAATTTATGATGGAAGATACCGGAGT 480  
Db 429 CAGATGTACTTTTCATTTGAAGAACTAATTAATAAATTTATGATGGAAGATACCGGAGT 488  
Qy 481 AGAGAAATGCGGTTTATTAATGATGAATTTTGTGAGTGTGGAATGCCCTTGCTGAA 540  
Db 489 AGAGAAATGCGGTTTATTAATGATGAATTTTGTGAGTGTGGAATGCCCTTGCTGAA 548  
Qy 541 TATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 549 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608  
Qy 601 AAAGATCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
Db 609 AAAGATCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668  
Qy 661 GATAAATTTTGAAGCAATTTCTCAATGTTTCCAGTAAAGGCGACAGCAAGAACTA 720  
Db 669 GATAAATTTTGAAGCAATTTCTCAATGTTTCCAGTAAAGGCGACAGCAAGAACTA 728  
Qy 721 AAGGAAAATTAAGAATCAACGAGAGAGCTCCAGGCGGACTTCTCTGAAATGT 780  
Db 729 AAGGAAAATTAAGAATCAACGAGAGAGCTCCAGGCGGACTTCTCTGAAATGT 788  
Qy 781 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 789 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848  
Qy 841 TTTTATGCGTTTCTGTAGGCGATGTTTAAATATGATGCTTCTTCAATCCTTTTCAAT 900  
Db 849 TTTTATGCGTTTCTGTAGGCGATGTTTAAATATGATGCTTCTTCAATCCTTTTCAAT 908  
Qy 901 GCAACACCAACATTAAGCGGAGAAACAGAAACAGCTCTAGACAAACAACTTGT 960  
Db 909 GCAACACCAACATTAAGCGGAGAAACAGAAACAGCTCTAGACAAACAACTTGT 968  
Qy 961 GGAACCAAGTGTACAGCAATTTGAGAGAGCAAAAGAGTTTGTGCTGCTCTCACGCT 1020  
Db 969 GGAACCAAGTGTACAGCAATTTGAGAGAGCAAAAGAGTTTGTGCTGCTCTCACGCT 1028  
Qy 1021 GAGCGATTAAGACCCACCAAAACGTCAGAGAGCGCGCAGAAAGAGAGCGCTTCCCAAT 1080

Dp	1029	GAGGGATTAAGACCCCAACAAAAGTCCAGAGAGGCGCAGAGAGGACGGCTTCCCAAT	1088
Qy	1081	AAACAGTAGCAGGCGCCAGACACCCCAACATTTAATGTGCTGAATCAAGAATACAGACGT	1140
Dp	1089	AAACAGTAGCAGGCCCAACGACCCCAACATTTAATGTGCTGAATCAAGAATACAGACGT	1148
Qy	1141	GATAGGGAAGCAGGGACTGAAAACGGGGGAGAGAACATGATTAAGAAAGAGAGAG	1200
Dp	1149	GATAGGGAAGCAGGGACTGAAAACGGGGGAGAGAACATGATTAAGAAAGAGAGAG	1208
Qy	1201	AAAGATGAATTTGAGAGTCTCTGGAAGCAATTTCTGGGTGCAAAACCAATTAAGATG	1260
Dp	1209	AAAGATGAATTTGAGAGTCTCTGGAAGCAATTTCTGGGTGCAAAACCAATTAAGATG	1268
Qy	1261	AAGCAAAATATTGAACCTCCTGAGAAATGTGAGTGGTGTGTAAGCCTCAATGTT	1320
Dp	1269	AAGCAAAATATTGAACCTCCTGAGAAATGTGAGTGGTGTGTAAGCCTCAATGTT	1328
Qy	1321	AGAGTCTCATTTGGCACTTACTATGACAATTTCTGTGCAATTEGTAGGTTAATTTGGAC	1380
Dp	1329	AGAGTCTCATTTGGCACTTACTATGACAATTTCTGTGCAATTEGTAGGTTAATTTGGAC	1388
Qy	1381	AAAACTGTACACAGGTGTATGAGTTTATGAGTCAAAAGATCTTACATCATAGCTCCAGCT	1440
Dp	1389	AAAACTGTACACAGGTGTATGAGTTTATGAGTCAAAAGATCTTACATCATAGCTCCAGCT	1448
Qy	1441	CCCGCTGAGAGTGTGATATCTCTCCAAAGAAAAGAAAGAAACACCGGTTGTGGCT	1500
Dp	1449	CCCGCTGAGAGTGTGATATCTCTCCAAAGAAAAGAAAGAAACACCGGTTGTGGCT	1508
Qy	1501	GCACATGTGAGAAAAGATACAGCTGAAAAAGACGGCTCTCTTAACATGTTTACAATAT	1560
Dp	1509	GCACATGTGAGAAAAGATACAGCTGAAAAAGACGGCTCTCTTAACATGTTTACAATAT	1568
Qy	1561	CAACCTGTGTATCATCCACGGCAGCCTTGTGTACAGTTGTGTGCTTGTGTGATGACCAA	1620
Dp	1569	CAACCTGTGTATCATCCACGGCAGCCTTGTGTACAGTTGTGTGCTTGTGTGATGACCAA	1628
Qy	1621	AAATTTTGTGAAAAGTTTGTCAATGTAGTTCAAGATGTCAAAAACCGCTTCCGGGATGC	1680
Dp	1629	AAATTTTGTGAAAAGTTTGTCAATGTAGTTCAAGATGTCAAAAACCGCTTCCGGGATGC	1688
Qy	1681	CGCTGCAAAAGCACAGTGTGACCAACCAAGCAGTGCCCGTGTCACTGGCTGTCCAGAGTGT	1740
Dp	1689	CGCTGCAAAAGCACAGTGTGACCAACCAAGCAGTGCCCGTGTCACTGGCTGTCCAGAGTGT	1748
Qy	1741	GACCTGTGACTGTGTCTTACTTGTGTGAGCCGCTGACCATTTGGACAGTAAAAATGTGTCC	1800
Dp	1749	GACCTGTGACTGTGTCTTACTTGTGTGAGCCGCTGACCATTTGGACAGTAAAAATGTGTCC	1808
Qy	1801	TGCAAGAACTGACAGTAATCAGCGGGGCTCCAAAAAGCATCTAATGCTGTGGCACATGTAC	1860
Dp	1809	TGCAAGAACTGACAGTAATCAGCGGGGCTCCAAAAAGCATCTAATGCTGTGGCACATGTAC	1868
Qy	1861	GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGCAAAAAATGAATTCATCTCAGAA	1920
Dp	1869	GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGCAAAAAATGAATTCATCTCAGAA	1928
Qy	1921	TACTGTGAGAGATTAATTTCTCAGATGAACTGTACAGAAAGGAAAAGTGTATGATTA	1980
Dp	1929	TACTGTGAGAGATTAATTTCTCAGATGAACTGTGTGAGTGTGAGTGTGATGATTA	1988
Qy	1981	TACATGTGACGTTTCTGTGTCAACTTGAACATGATTTTGTGTGAGTGTGAGTGTGAG	2040
Dp	1989	TACATGTGACGTTTCTGTGTCAACTTGAACATGATTTTGTGTGAGTGTGAGTGTGAG	2048
Qy	2041	GGTATCAAAATTCGTTTTGGCAAAATCATTCGGTAAATCCAACTGTATGCAAAAGTTATG	2100
Dp	2049	GGTATCAAAATTCGTTTTGGCAAAATCATTCGGTAAATCCAACTGTATGCAAAAGTTATG	2108
Qy	2101	ATGTGTAAACGTGATCACAGATAGTATTTTGTCCAAAGAGCCATTCAGACTGGCGAA	2160

DB	2109	ATGTTAACGGGTATCA CAGGATAGTATATTTTCCAAAGAGACATTCAGACTGGCGAA	2168
QY	2161	GAGCTGTTTTTGTGATTACAGATACAGCCAGGCTGATCCCTCGAAGTATGTCGGATCGAA	2220
DB	2169	GAGCTGTTTTTGTGATTACAGATACAGCCAGGCTGATCCCTCGAAGTATGTCGGATCGAA	2228
QY	2221	AGAGAAATGGAATCCCTTGA	2241
DB	2229	AGAGAAATGGAATCCCTTGA	2249
RESULT 15			
ADP95922	ID	ADP95922 standard; DNA; 2695 BP.	
ADP95922;			
07-OCT-2004	(first entry)		
Human EZH2	(enhancer of zeste homologue 2)-encoding DNA, SEQ ID NO.3.		
Human; EZH2;	enhancer of zeste homologue 2; chromosome 7q35;		
histone methyltransferase; histone H3; cell cycle; tumorigenesis;			
inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;			
lung cancer; pancreatic cancer; ovarian cancer; drug screening;			
prophylaxis; prevention; diagnosis; cytostatic; gene; de.			
Homo sapiens.			
Key	Location/Qualifiers		
CD5	167..2422		
FT	/*tag= a		
FT	/product= "EZH2 (enhancer of zeste homologue 2) "		
PN	WO2004058969-A1.		
XX	15-JUL-2004.		
PF	22-DEC-2003; 2003WO-JP016417.		
PR	24-DEC-2002; 2002JP-00373144.		
PA	(TAKE ) TAKEDA CHEM IND LTD.		
PI	Hikichi Y, Nishizawa S;		
DR	WPI: 2004-525885/50.		
DR	P-PsDB; ADP95920.		
PT	Novel prophylactic and therapeutic agent of cancer, having compound or		
PT	its salt that inhibits activity of enhancer of zeste Drosophila homologue2		
PT	protein, useful for treating cancer such as lung cancer, breast cancer.		
XX	Example 3; SEQ ID NO 3; 102pp; Japanese.		
CC	The invention relates to an agent for the prophylaxis and therapy of		
CC	cancer comprising a compound which inhibits the activity or expression of		
CC	human EZH2 (enhancer of zeste homologue 2), resulting in apoptosis. The		
CC	EZH2 inhibitors include anti-EZH2 antibodies and EZH2 antisense nucleic		
CC	acids. The invention also relates to a kit for screening for inhibitors		
CC	of EZH2 activity or expression. EZH2 is a component of a complex with		
CC	histone methyltransferase activity which methylates histone H3 on lysine		
CC	residues at position 9 and/or 27. It has been found to be deregulated in		
CC	various tumours, and acts as a mediator of tumorigenesis, being able to		
CC	reverse repression of cyclin A which in turn mediates cell cycle		
CC	advancement. The agent of the invention can be used for preventing,		
CC	treating and diagnosing cancers such as colorectal cancer, breast cancer,		
CC	lung cancer, pancreatic cancer and ovarian cancer, and can also be used		
CC	in drug screening. The agent of the invention is safe and less toxic than		
CC	prior art cancer prophylactic/therapeutic agents. The present sequence		
CC	represents a DNA encoding human EZH2 which was used in an example of the		
CC	invention.		



SQ Sequence 2695 BP; 839 A; 549 C; 660 G; 647 T; 0 U; 0 Other;

Query Match 98.9%; Score 2216; DB 12; Length 2695;  
Beet Local Similarity 99.3%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 ATGGCCAGACTGGGAGAAATCTGAGAGGAGCCAGTTGTTGGCGAAGGCTGTAA 60  
DB 167 ATGGCCAGACTGGGAGAAATCTGAGAGGAGCCAGTTGTTGGCGAAGGCTGTAA 226  
QY 61 TCAGATCATGCGACTGAGACGCTCAAGAGTTTCAAGAGCTGATGATGAAAGT 120  
DB 227 TCAGATCATGCGACTGAGACGCTCAAGAGTTTCAAGAGCTGATGATGAAAGT 286  
QY 121 ATGTTAGTTCCAAATGCTGAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATGG 180  
DB 287 ATGTTAGTTCCAAATGCTGAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATGG 346  
QY 181 AAACGCGAAGGATCAGCCTGTCACATCTGACTTCTGTAGGCTCATTTGGCGGAGT 240  
DB 347 AAACGCGAAGGATCAGCCTGTCACATCTGACTTCTGTAGGCTCATTTGGCGGAGT 406  
QY 241 AGGAGTGTTCGGTGAACAGTGAATTTTCCAAACAAGTCAATCCCATTAAGAAT 300  
DB 407 AGGAGTGTTCGGTGAACAGTGAATTTTCCAAACAAGTCAATCCCATTAAGAAT 466  
QY 301 CTGAATGCAATGCTTCAATACATTAATGTTCTTGGTCTCCCTTACAGCAAAATTT 360  
DB 467 CTGAATGCAATGCTTCAATACATTAATGTTCTTGGTCTCCCTTACAGCAAAATTT 526  
QY 361 ATGTGGAAGATGAATGTTTACATTAATCTCTATATGGAAGATGATGATTTAGT 420  
DB 527 ATGTGGAAGATGAATGTTTACATTAATCTCTATATGGAAGATGATGATTTAGT 586  
QY 421 CAGATGATCTTCTTATGAGAACTAATATTAATATGATGGAAGATGACCGGAGT 480  
DB 587 CAGATGATCTTCTTATGAGAACTAATATTAATATGATGGAAGATGACCGGAGT 646  
QY 481 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGTTGTGATGCTTGTGTCAA 540  
DB 647 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGTTGTGATGCTTGTGTCAA 706  
QY 541 TATATGATGATGATGATGATGATGAGACGATCTCGAAGAAAGAAAGAAAGAG 600  
DB 707 TATATGATGATGATGATGATGATGAGACGATCTCGAAGAAAGAAAGAAAGAG 766  
QY 601 AAAGATCTGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTT 660  
DB 767 AAAGATCTGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTT 826  
QY 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAACAGCAAGAACTA 720  
DB 827 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAACAGCAAGAACTA 886  
QY 721 AAAGAAATTAAGAACTACCGAGACGACGCTCCAGGGGCACTTCTCTGTAAGT 780  
DB 887 AAAGAAATTAAGAACTACCGAGACGACGCTCCAGGGGCACTTCTCTGTAAGT 946  
QY 781 ACCCCCAATGATGACCAATGCTAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840  
DB 947 ACCCCCAATGATGACCAATGCTAATCTGTTCAAGAGAGCAAAAGCTTACATCC 1006  
QY 841 TTTCAATGCTTTCTGTAGCGATGTTTAAATATGATGCTTCTTACATC----- 892  
DB 1007 TTTCAATGCTTTCTGTAGCGATGTTTAAATATGATGCTTCTTACATGTAAGTC 1066  
QY 893 -----CTTTTCAATGACCAACCACTTATTAAGCGGAGAGAAACAAGAAACACTCTA 945  
DB 1067 AATTATTTCTTTTCAATGACCAACCACTTATTAAGCGGAGAGAAACAAGAAACACTCTA 1126  
QY 946 GACAAACAACCTTGTGACCAACGTTTACAGCAATTTGAGGGAGCAAGAGATTTCT 1005  
DB 1127 GACAAACAACCTTGTGACCAACGTTTACAGCAATTTGAGGGAGCAAGAGATTTCT 1186

QY 1006 GCTGCTCACCGCTGAGCGGATTAAGACCCCAACAAAACGTCAGAGGCGCGAGAGA 1065  
DB 1187 GCTGCTCACCGCTGAGCGGATTAAGACCCCAACAAAACGTCAGAGGCGCGAGAGA 1246  
QY 1066 GAGCGGCTTCCCAATTAAGATGACAGGCGCCAGACCCCCACCAATTAATGTGTGATCA 1125  
DB 1247 GAGCGGCTTCCCAATTAAGATGACAGGCGCCAGACCCCCACCAATTAATGTGTGATCA 1306  
QY 1126 AAGATTAACAACGATTAAGAGAGCAAGGAGCTGAACCGGGGGAGAGAACATGATAA 1185  
DB 1307 AAGATTAACAACGATTAAGAGAGCAAGGAGCTGAACCGGGGGAGAGAACATGATAA 1366  
QY 1186 GAAGAAAG 1245  
DB 1367 GAAGAAAG 1426  
QY 1246 ACACAATTAAGATGAAGCCAAATTAAGACCTCTGAGAAATGTGAGTGAAGTGTCT 1305  
DB 1427 ACACAATTAAGATGAAGCCAAATTAAGACCTCTGAGAAATGTGAGTGAAGTGTCT 1486  
QY 1306 GAAGCTCAATGTTTGAAGTCTCAATTTGCACTTAATGACATTTCTGTGCAATGCT 1365  
DB 1487 GAAGCTCAATGTTTGAAGTCTCAATTTGCACTTAATGACATTTCTGTGCAATGCT 1546  
QY 1366 AGCTTAATTTGGAACCAAAACATGTAAGAGGATGATGATTTAGAGTCAAGAACTAGC 1425  
DB 1547 AGCTTAATTTGGAACCAAAACATGTAAGAGGATGATGATTTAGAGTCAAGAACTAGC 1606  
QY 1426 ATCAATGCTCAGCTCCCGTGAAGATGATGATCTCTCCAGAGAAAGAAAGAGAGAA 1485  
DB 1607 ATCAATGCTCAGCTCCCGTGAAGATGATGATCTCTCCAGAGAAAGAAAGAGAGAA 1666  
QY 1486 CACCGTGTGGGCTGCACTGCAAGAAATACGCTGAAGAAAGAGAGAGAGAGAGAG 1545  
DB 1667 CACCGTGTGGGCTGCACTGCAAGAAATACGCTGAAGAAAGAGAGAGAGAGAGAG 1726  
QY 1546 CATGTTTCAACTATCAACCTGTGATATCAACGAGGCGCTTGTGACATGTTCTGTCT 1605  
DB 1727 CATGTTTCAACTATCAACCTGTGATATCAACGAGGCGCTTGTGACATGTTCTGTCT 1786  
QY 1606 TGTGATGATGACAAATTTTGTGAAAGGTTTGTCAATGATGCTCAGAGTCAAAAC 1665  
DB 1787 TGTGATGATGACAAATTTTGTGAAAGGTTTGTCAATGATGCTCAGAGTCAAAAC 1846  
QY 1666 CGCTTTCGGGATGCGCTGCAAGACAGTGAACACCAAGAGTGCCTGTACTCTG 1725  
DB 1847 CGCTTTCGGGATGCGCTGCAAGACAGTGAACACCAAGAGTGCCTGTACTCTG 1906  
QY 1726 GCTGTCCGAGATGACCTGACCTGTCTTACTTGTGAGGCGCTGACATTTGGGAC 1785  
DB 1907 GCTGTCCGAGATGACCTGACCTGTCTTACTTGTGAGGCGCTGACATTTGGGAC 1966  
QY 1786 AGTAAATATGTCTCTGCAAGAACGTCAGTATCAAGCGGGGCTCAAAAGATCTATTG 1845  
DB 1967 AGTAAATATGTCTCTGCAAGAACGTCAGTATCAAGCGGGGCTCAAAAGATCTATTG 2026  
QY 1846 CTGGCAACATCTGACGTGCAAGGCTGGGGATTTTATCAAGATCTCTGTCAAGAAAT 1905  
DB 2027 CTGGCAACATCTGACGTGCAAGGCTGGGGATTTTATCAAGATCTCTGTCAAGAAAT 2086  
QY 1906 GAATTCATCTCAAGATTAATGAGAGATTTTCTCAAGATGAAGCTGACAGAGAGG 1965  
DB 2087 GAATTCATCTCAAGATTAATGAGAGATTTTCTCAAGATGAAGCTGACAGAGAGG 2146  
QY 1966 AAGGTATGATTAATATCATGTCAGGCTTCTGTCAACTTGAACAATGATTTTGTGTG 2025  
DB 2147 AAGGTATGATTAATATCATGTCAGGCTTCTGTCAACTTGAACAATGATTTTGTGTG 2206  
QY 2026 GATCAACCCGCAAGGTTAACAATTCGTTTGCATTAATTCGTTAATTCAAATCTGC 2085  
DB 2207 GATCAACCCGCAAGGTTAACAATTCGTTTGCATTAATTCGTTAATTCAAATCTGC 2266

Qy	2086	TATGCAAAAGTTATGATGGTTAACGTCATCAGATAGGTATTTTGCCAGAGAGCC	2145
Db	2267	TATGCAAAAGTTATGATGGTTAACGTCATCAGATAGGTATTTTGCCAGAGAGCC	2326
Qy	2146	ATCGAGACTGGCCGAAAGCTGTTTTTGATTACAGATACAGCCAGGCTGATGCCCTGAAG	2205
Db	2327	ATCGAGACTGGCCGAAAGCTGTTTTTGATTACAGATACAGCCAGGCTGATGCCCTGAAG	2386
Qy	2206	TATGTCGGCATCGAAAGAAATGGAATCCCTTGA	2241
Db	2387	TATGTCGGCATCGAAAGAAATGGAATCCCTTGA	2422

Search completed: August 13, 2006, 17:13:56  
 Job time : 1419 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2006, 11:16:17 ; Search time 10754 Seconds  
(without alignments)  
11652.894 Million cell updates/sec

Title: US-10-773-302-1\_COPY\_90\_2330  
Perfect score: 2241  
Sequence: 1 atggcgccgactgggaagaa.....gagaatggaatcccttga 2241

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_hic: \*  
7: gb\_est2: \*  
8: gb\_est7: \*  
9: gb\_est8: \*  
10: gb\_est9: \*  
11: gb\_g881: \*  
12: gb\_g882: \*  
13: gb\_g883: \*  
14: gb\_g884: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1970.6	87.9	2652	6	AK157853 Mus muscu
2	1866	83.3	2241	14	AY399441 Homo sapi
3	1654.8	73.8	2241	14	AY399443 Mus muscu
4	1590.2	71.0	2241	14	AY399442 Pan trogl
5	1017.8	45.4	1238	6	AK153835 Mus muscu
6	871.4	38.7	873	8	CR990276 CR990276
7	868.2	38.9	967	3	BQ946546 AGENCOURT
8	848.4	37.9	956	3	BQ653044 AGENCOURT
9	846.2	37.8	972	3	BQ656770 AGENCOURT
10	840.4	37.5	842	9	CR870387 HESCA_44
11	832.4	37.1	4643	6	CR857825 Pongo pyg
12	830	37.0	837	10	DR762983 HESCA_147
13	828.8	37.0	4089	6	AK154565 Mus muscu
14	828.8	37.0	4194	6	AK140694 Mus muscu
15	826.4	36.9	840	9	CX783825 HESCA_26
16	826.4	36.9	1019	3	BQ65838 AGENCOURT
17	821.2	36.6	901	4	BK459867 AGENCOURT
18	811.8	36.2	971	8	CO647894 ILUMIGEN
19	802.4	35.8	2244	14	AY402396 Homo sapi

20	792	35.3	965	3	BQ654181 AGENCOURT
21	782.2	34.9	2215	14	AY402398
22	771.6	34.4	946	7	BE729850
23	769.6	34.3	876	2	BM040808
24	769.6	34.3	909	3	BQ165381 AGENCOURT
25	768.6	34.3	967	3	BQ649256 AGENCOURT
26	765.6	34.2	904	4	CA454418 AGENCOURT
27	761.4	34.0	870	4	BX462271
28	756.2	33.7	825	1	AU131868
29	751.8	33.5	775	2	B1858128
30	740	33.0	740	9	CR872564
31	738	32.9	2183	14	AY402397
32	734.2	32.8	859	4	BM974704
33	730.4	32.6	879	4	CR983448
34	725.4	32.4	894	8	CR983448
35	725	32.4	813	1	AU123971
36	724	32.3	804	2	BM456354
37	718.4	32.1	933	10	DV806568
38	718.2	32.0	974	7	BF026728
39	716.4	32.0	869	10	DV927839
40	714.6	31.9	833	1	AU124471
41	714.6	31.9	920	3	BQ508767
42	713.2	31.8	1071	2	BQ393349
43	712.2	31.8	893	3	BQ189096
44	708.4	31.6	878	3	BQ184475
45	706	31.5	859	10	DT722075

#### ALIGNMENTS

RESULT 1	AK157853	2652 bp	mRNA	linear	HTC 21-SEP-2005
LOCUS	AK157853				
DEFINITION	Mus musculus erythroblast cDNA, RIKEN full-length enriched library, clone:K0C0026B09 product:enhancer of zeste homolog 2 (Drosophila), full insert sequence.				
ACCESSION	AK157853	GI:74186125			
VERSION	AK157853.1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
REFERENCE	10349636				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
REFERENCE	11042159				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
REFERENCE	11076861				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,				

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
11217851

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, T.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,  
Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L. E., Cousine, S., Dalla, E., Dragani, T. A.,  
Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
Gardibidi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, B. D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. Z., Kongaya, A.,  
Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
Maltais, L., Marchionni, L., McKenrie, L., Miki, H., Nagashima, T.,  
Nunata, K., Okido, T., Pavan, W. J., Perera, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L. G., Wyszewski, B. S., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carinici, P.,  
Hayakawa, N., Hirozane-Kishikawa, T., Komno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imochi, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Saeki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)  
12466851

Carrini, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C.,  
Macedo, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Koduru, R.,  
Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R.,  
Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E.,  
Amesbury, M., Baxendale, A., Apweiler, R., Attaluri, R. N., Bailey, T. L.,  
Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chak, A. M.,  
Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R.,  
Crome, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G.,  
di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Paulner, G.,  
Fletcher, C. F., Fukushima, T., Furuno, M., Fukui, S., Gardibidi, M.,  
Georgil-Hemmings, P., Gingeras, T. R., Gojobori, T., Green, R. E.,  
Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N.,  
Hill, D., Hummel, L., Iacono, M., Ikey, K., Iwama, A., Ishikawa, H.,  
Jait, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kato, J., Kikuchi, S. K.,  
Kikuchi, H., Kollas, G., Krieban, S. P., Kruger, A., Kummerow, S. K.,  
Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J.,  
Liu, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L.,  
Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S.,

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)  
16141072

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

RIKEN Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)  
16141073

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

8 (bases 1 to 2652)  
Arakawa, T., Carrini, P., Fukuda, S., Hashizume, W., Hayashida, K.,  
Hori, P., Iida, Y., Imanura, K., Imotani, K., Itoh, M., Kanagawa, S.,  
Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Niimiya, N.,  
Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Saeki, D.,  
Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watabiki, A.,  
Miyazawa, M. and Hayashizaki, Y.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

ENCODE  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Tissues were provided by Hiromitsu Nakauchi (Center for  
Experimental Medicine, Institute of Medical Science, University of  
Tokyo 4-6-1, Shirokane-dai, Minato-ku, Tokyo, 108-8639, JAPAN )  
whose assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

FEATURES  
source  
1. 2652  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM:DB:K0C0026B09"  
/db\_xref="taxon:10090"  
/clone="K0C0026B09"  
/cell\_type="erythroblast"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
146. 2386

CDS

/note="unnamed protein product; enhancer of zeste homolog  
2 (Drosophila) (MGD|MGI:107940 GB|NM\_007971, evidence:  
BLASTN, 99%, match=2635)  
putative"

/codon\_start=1  
/protein\_id="BAE34232.1"  
/db\_xref="gi:74186126"

/translation="MGQTKSKKGPVCKRKYSEWRLQARPRADRYKMRSS  
NRQKLEBETLNQKQRIQHIHTSVSLRSTRCSVTSDDLPQVPLKTLN  
AVASVPIWSPLOQNFVDETVLNIPIWGDVLDQDTPLELKNYDGVHD  
RBCGFINDEIVELVNALGVNDDDDDGDDPDERBKQDLBNRDKICPRK  
PADKIFEAISWFPKGTABELKEKELTEQOLGALPPECTPIIDPKMSVREQ  
SLHSFTLCRRRCFYDCETLHPFHAIPMYTKRMTETLDMKPCGPOCYQLBQKRF  
AAALAEIKTPKRRGRRLPNNSRSTPTISVLESMDIDSDBAEHTGGEN

Query Match 87.9%; Score 1970.6; DB 6; Length 2652;

Best Local Similarity 92.5%; Pred. No. 0;

Matches 2072; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
QY 1 ATGGCCGAGCTGGGAAGAAATCTGAGAAAGGACCAAGTTGTGGCGGAAGCGTGTAA 60
DB 146 ATGGCCGAGCTGGGAAGAAATCTGAGAAAGGACCGGTTGTGGCGGAAGCGTGTAA 205
QY 61 TCAGAGTACATGCGACGTGACAGCTCAAGAGTTCAAGCAGATGAGTAAGTAAGAGT 120
DB 206 TCAGAGTACATGAGCTGAGACGCTCAAGAGTTCAAGAGCTGATGAGTAAGTAAGAGT 265
QY 121 ATGTTAGTCCATGCTGCAAAAATTTTGGAAAGACGGAATTTTAAACCAAGATGG 180
DB 266 ATGTTAGTCCATGCTGCAAAAATTTTGGAAAGACGGAATTTTAAACCAAGATGG 325
QY 181 AAACGCGGAAGATCAGCCTGTGACATCTGACTTCTGTAGCTCATTTGGCGGAGCT 240
DB 326 AAACGCGGAAGATCAGCAGTGTGACATCATGACTTCTGTAGCTCATTTGGCGGAGCT 385
QY 241 AGGAGTGTTCGATGACAGTGTGATTTTCCAAACAGATCATCCCATTTAAAGCT 300
DB 386 AGGAGTGTTCAGTACAGTGTGATTTTCCAAACAGATCATCCCATTTAAAGCT 445
QY 301 CTGAATGAGTGTCTTCAAGTACCAATATGATTTCTGGTCTCCCTTACAGAGATTTT 360
DB 446 CTGAATGAGTGTCTTCAAGTACCAATATGATTTCTGGTCTCCCTTACAGAGATTTT 505
QY 361 ATGCTGGAAGATGAACTGTTTACATTAACATTTCTTATATGAGGAGATGAAAGTTTAAT 420
DB 506 ATGCTGGAAGATGAACTGTTTACATTAACATTTCTTATATGAGGAGATGAAAGTTTCTGAT 565
QY 421 CAGATGTGATCTTCAATGAGAACTAATTAATAAATTTATGATGGAAAGTACACGGGAT 480
DB 566 CAGATGTGATCTTCAATGAGAACTAATTAATAAATTTATGATGGAAAGTACATGATGAC 625
QY 481 AGAGATGTGGTGTATTAATGATGAAATTTTGTGAGTGGTGTGATGCTTGTGTCA 540
DB 626 AGAGATGTGGTGTATTAATGATGAAATTTTGTGAGTGGTGTGATGCTTGTGTCA 685
QY 541 TTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 686 TTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 745
QY 601 AAAGATCTGAGAGATCAGCAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660
DB 746 AAAGATCTGAGAGATCAGCAGATGATTAAGAAAGCTTCCCACTCGGAAATTTCTTCT 805
QY 661 GATAAATATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAGAAAGTA 720
DB 806 GATAAATATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAGAAAGTA 865
QY 721 AAAGAAATATTAAGAACTACCGAAGCAGCTCCAGGCGCAGCTTCTCTGAAATGT 780
DB 866 AAAGAAATATTAAGAACTACCGAAGCAGCTCCAGGCGCAGCTTCTCTGAAATGT 925
QY 781 ACCCCCAATAGATGACAAATGCTAAATCTGTTGAGAGAGCAAGGCTTACACTCC 840
DB 926 ACTCCAAATCATGATGACAAATGCTAAATCTGTTGAGAGAGCAAGGCTTACACTCC 885
QY 841 TTTCAATAGCTTTTCTGATGAGGATGTTTAAATATGATCTGCTTCTTCAATCTTTTCA 900
DB 986 TTTCAATAGCTTTTCTGATGAGGATGTTTAAATATGATCTGCTTCTTCAATCTTTTCA 1045
QY 901 GCAACACCCCAATCTTATAGCGGAGAAACAGAAACAGCTCTGAGCAACAACTTGT 960
DB 1046 GCAACACCCCAATCTTATAGCGGAGAAACAGAAACAGCTTGTGACAAACAGCTTGT 1105
QY 961 GACCAACAGTGTACCAAGATTTGAGAGGAGCAAGAGTGTGCTGCTCTCAACGCT 1020
DB 1106 GACCAACAGTGTACCAAGATTTGAGAGGAGCAAGAGTGTGCTGCTCTCTCAACGCT 1165
QY 1021 GAGCGATTAAGAACCCCAACCAAACTCAGAGAGGCGCCAGAAAGAGCGCTTCCCAT 1080
DB 1166 GAGCGATTAAGAACCCCAACCAAACTCAGAGAGGCGCCAGAAAGAGCGCTTCCCAT 1225
QY 1081 AACAGTACAGAGCCAGAGACCCCACTTATATGCTGGAATCAAGATACAGACAGT 1140
DB 1226 AACAGTACAGAGCCAGAGACCCCACTTATATGCTGGAATCAAGATACAGACAGT 1285
QY 1141 GATAGGAGAGAGGAGCTGAAACGAGGAGAGAGAACTAATTAAGAAAGAGAGAG 1200
DB 1286 GATAGGAGAGAGGAGCTGAAACGAGGAGAGAGAACTAATTAAGAAAGAGAGAGAG 1345
QY 1201 AAAGTGAATCTTGAAGCTCTCTGAGCAAAATTTCTGGTGTCAAAACCAATTAAGATG 1260
DB 1346 AAAGTGAAGAGTCAAGCTCTCTGAGCAAAATTTCTGGTGTCAAAACCAATTAAGATG 1405
QY 1261 AAAGCAATATTAAGACCTCCTGAGAAATGAGAGTGGAGTGGAGCTCAATGTTT 1320
DB 1406 AAAGCAATATTAAGACCTCCTGAGAAATGAGAGTGGAGTGGAGCTCAATGTTT 1465
QY 1321 AGAGTCTCATTTGAGCACTTATAGCAATTTCTGTGCAATTTCTAGGTTAATGGAGC 1380
DB 1466 AGAGTCTCATTTGAGCACTTATAGCAATTTCTGTGCAATTTCTAGGTTAATGGAGC 1525
QY 1381 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAAATTTAGCATATATGCTCAGCT 1440
DB 1526 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAAATTTAGCATATATGCTCAGCT 1585
QY 1441 CCGCGTGAAGATGAGATCACTCTCAAGAGAAAGAGAGAAACAGCGTGTGGGCT 1500
DB 1586 CCGCGTGAAGATGAGATCACTCTCAAGAGAAAGAGAGAAACAGCGTGTGGGCT 1645
QY 1501 GCACATGTCAGAAAGATCAGCTGAAAAAGAGACGCTCTCTTAACATATTTTCAACTAT 1560
DB 1646 GCACATGTCAGAAAGATCAGCTGAAAAAGAGACGCTCTCTTAACATATTTTCAACTAT 1705
QY 1561 CAACCTGTGTATCAACAGGAGCTTGTGACAGTTCGCTTGTGTATGACAA 1620
DB 1706 CAACCTGTGTATCAACAGGAGCTTGTGACAGTTCGCTTGTGTATGACAA 1765
QY 1621 AATTTTGTGAAAGTTTGTCAATGTATGATCAAGTGTCAAAACCGCTTCCGGGATGC 1680
DB 1766 AATTTTGTGAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGAT 1825
QY 1681 CGCTGCAAGAGCAGTGTGACCAAGCAGTCCGCTGTCACTGTGCTGTCCGAGAGTGT 1740
DB 1826 CGCTGCAAGAGCAGTGTGACCAAGCAGTCCGCTGTCACTGTGCTGTCCGAGAGTGT 1885
QY 1741 GACCTGTGACTGTGTCTTATCTTGTGAGCGGCTGTACCAATGGAGCAGTAAATGTGTCC 1800
DB 1886 GACCTGTGACTGTGTCTTATCTTGTGAGCGGCTGTACCAATGGAGCAGTAAATGTGTCC 1945
QY 1801 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAGAGATCTAATGCTGTGGCAGCATGTGAC 1860
DB 1946 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAGAGATCTAATGCTGTGGCAGCATGTGAC 2005
QY 1861 GTGCAAGGCTGGGAGATTTTATCAAAAGATCTGTGAGAAATTAATGATCTCAAGAA 1920
DB 2006 GTGCAAGGCTGGGAGATTTTATCAAAAGATCTGTGAGAAATTAATGATCTCAAGAA 2065
```

QY	1921	TACGTGTGAGAGATTAATTCTTCACATAGAGCTGCACAAGAGGAAAGTGTATGATPAA	1980
Db	2066	TACTGTGGGAGATTAATTTCTCAGATGAGCAGACAGAAAGGAAAGTGTATGACAA	2125
QY	1981	TACATGTGAGGCTTTCGTGTCAACTTGAACATGATTTTGTGTGTGATGCACCCGACAG	2040
Db	2126	TACATGTGAGGCTTTCGTGTCAACTTGAACATGATTTTGTGTGTGATGCACCCGAAAG	2185
QY	2041	GGTAAACAAATTCGTTTGGCAATTCATTGCGTAAATCCAACTGCTATGCAAAAGTTATG	2100
Db	2186	GGCAACAAATTCGTTTGGCTATATTCATTCAGTAAATCCAAACTGCTATGCAAAAGTTATG	2245
QY	2101	ATGTGTAACGGTATCACAGATAGATGATTTTGTGGCAGAGAGGACATGCAGCTGGCCGA	2160
Db	2246	ATGTGTAATGTGACCAAGATAGGACATCTTGTCTAAGAGGCTATTCACATGTGTGAA	2305
QY	2161	GAGCTGTATTTTGTATTACAGATACAGACGAGCTGATGCGCTGAAGTATGTGCGCATCGAA	2220
Db	2306	GAGTTGTTTGTATTACAGATACAGCGCAGGCTGATGCGCTGAAGTATGTGCGCATCGAA	2365
QY	2221	AGAGAAATGGAATCCCTTGA	2241
Db	2366	CGAGAAATGGAATCCCTTGA	2386

RESULT 2

AY399441

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

COMMENT

FEATURES

ORIGIN

Query Match

Best Local Similarity

Matches 1866;

Conservative

83.3%;

Score 1866;

DB 14;

Length 2241;

Pred. No. 0;

Mismatches 375;

Indels 0;

Gaps 0;

1

ATGGGCGCAGACTGGAGAAATCTGAGAGGAGCAGATTGTGTGGCGGAGAGGTGTAAAA

60

Db 1 ATGGGCCAGCTGGGAAAGAAATCTGAAAGGAGACAGATTGTTGGCGAAAGCCTGTAA 60  
QY TCAGAGTACATGCGACATCGAGCAGCTCAGAGGTTTCAGACAGCTGA TGAAGTAAAGT 120  
Db TCAGAGTACATGCGACATCGAGCAGCTCAGAGGTTTCAGACAGCTGA TGAAGTAAAGT 120  
QY ATGTTTAGTTCGAATCGTCAAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180  
Db ATGTTTAGTTCGAATCGTCAAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180  
QY AAACGAGGAGGATACAGCCTGTGACATCTGTCTCTGTGAGCTCATTTGGCGGGAGCT 240  
Db AAACGAGGAGGATACAGCCTGTGACATCTGTCTCTGTGAGCTCATTTGGCGGGAGCT 240  
QY AGGAGGTGTCGCTGACCGAGTCACTTGATTTTCCACACAGTCAATCCATTAAAGACT 300  
Db AGGAGGTT 300  
QY AGGAGGTT 300  
Db CTGAATGACGTTGCTTCAGTACCCATTAATGATTTCTTGTCCTCCCTACAGAGAAATTT 360  
QY NNN 360  
Db ATGTGGAAGATGAACCTGTTTACATPACCTCTATATGGAGATGAAGTTTATGAT 420  
QY NNNNGGAGAGTGAACCTGTTTACATPACCTCTATATGGAGATGAAGTTTATGAT 420  
Db CAGATGCTGATCTTCAATTGAAGATCATATAAAATTTATGATGGAAATGATCAAGGGAT 480  
QY AGAGATGTGGGTTTATATATGATGAATTTTGTGGAGTTGGTGAATGCCCTTGGTCA 540  
Db AGAGATGTGGGTTTATATATGATGAATTTTGTGGAGTTGGTGAATGCCCTTGGTCA 540  
QY TATATATGATGATGCGATGATGATGAGAGAGATCTGTAAGAAAGAGAAAGACAG 600  
Db TATATATGATGATGCGATGATGATGAGAGAGATCTGTAAGAAAGAGAAAGACAG 600  
QY AAAGTCTGAGATGACCGAGATGATPAAAGAAAGCCGCCACTCGGAAATTTCTTCT 660  
Db AAAGTCTGAGATGACCGAGATGATPAAAGAAAGCCGCCACTCGGAAATTTCTTCT 660  
QY GATTAATTTTGAAGCATTCTCTCATGTTTCCATATAGGGCACAAGCAAGAAACTA 720  
Db GATTAATTTTGAAGCATTCTCTCATGTTTCCATATAGGGCACAAGCAAGAAACTA 720  
QY AAGGAAATTTTAAGAACTCACGGAACAGAGCTCCAGGCGCACTTCCCTCGAATGT 780  
Db AAGGAAATTTTAAGAACTCACGGAACAGAGCTCCAGGCGCACTTCCCTCGAATGT 780  
QY ACCCCACATAGATGACCAATATGCTAAATCTGTTCAAGAGAGCAAGCTTACACTCC 840  
Db ACCCCACATAGATGACCAATATGCTAAATCTGTTCAAGAGAGCAAGCTTACACTCC 840  
QY TTTCAATCGCTTTTCTGTAGGCGATGTTTTAATATATGATGCTTCTTCAATCCTTTTCA 900  
Db TTTCAATCGCTTTTCTGTAGGCGATGTTTTAATATATGATGCTTCTTCAATCCTTTTCA 900  
QY GCAACACCCACACTTATTAACGGAGAAACACAAACAGCTTACACAAACCTTGT 960  
Db GCAACACCCACACTTATTAACGGAGAAACACAAACAGCTTACACAAACCTTGT 960  
QY NNN 960  
Db GAAACACAGTGTATACAGACTTTGTAGAGAGCAAGAGTTTGTGCTGCTCACCGCT 1020  
QY NNN 1020  
Db GAGCGATTAAGACCCCAACAAACGTCACAGAGGCGCAGAAAGAGACGGCTTCCCAAT 1080  
QY AACGATGACAGCCCAAGACCCCAACCAATTAATGTCTGAAATTAAGGATACAGACGT 1140

1081 AACAGTAGAGGCGCCAGCACCCACCATTAATGTGCTGGAATCAAAAGATACAGACGT 1140  
QY 1141 GATTAGGAAGCGAGCTGTAACCGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1200  
DB 1141 GATTAGGAAGCGAGCTGTAACCGGGGGAGAGAACATGATTAAGAAAGAGAGAG 1200  
QY 1201 AAAGATGAACCTTCGAGCTCTCTGAAAGAAAATTCGCGTGTCAAAACCAATTAAGATG 1260  
DB 1201 AAAGATGAACCTTCGAGCTCTCTGAAAGAAAATTCGCGTGTCAAAACCAATTAAGATG 1260  
QY 1261 AAGCCAAATATTAATTAACCTCTGGAAGATGTGAGTGAAGTGTGCTGAAGCTCAATGTTT 1320  
DB 1261 AAGCCAAATATTAATTAACCTCTGGAAGATGTGAGTGAAGTGTGCTGAAGCTCAATGTTT 1320  
QY 1321 AAGATCTCAATTTGGCCTTAATGACATTTCTGTGCCAATGCTAGTAAATTTGGAGAC 1380  
DB 1321 AAGATCTCAATTTGGCCTTAATGACATTTCTGTGCCAATGCTAGTAAATTTGGAGAC 1380  
QY 1381 AAAACATGTGACAGGTGTATGAGTTAGATCAAAAGATCTAGCATCATAGCTCAGCT 1440  
DB 1381 AAAACATGTGACAGGTGTATGAGTTAGATCAAAAGATCTAGCATCATAGCTCAGCT 1440  
QY 1441 CCGCGTAGAGATGTGATATCTCTCCAGAGAAAAGAGAGAAAACCGGTTGTGGCT 1500  
DB 1441 CCGCGTAGAGATGTGATATCTCTCCAGAGAAAAGAGAGAAAACCGGTTGTGGCT 1500  
QY 1501 GCAACTGTGCAAAAGATACAGCTGAAAAGACGCGCTCTCTAACCATGTTTCAACTAT 1560  
DB 1501 GCAACTGTGCAAAAGATACAGCTGAAAAGACGCGCTCTCTAACCATGTTTCAACTAT 1560  
QY 1561 CAACCTGTGATCATCCAGGAGCGCTTGTGACAGTCCGTCCTGTGATGATAGACAA 1620  
DB 1561 CAACCTGTGATCATCCAGGAGCGCTTGTGACAGTCCGTCCTGTGATGATAGACAA 1620  
QY 1621 AATTTTGTGAAAAGTTTGTCAATGTAGTTCAGATGTCAAAACCGCTTCCGGATGC 1680  
DB 1621 AATTTTGTGAAAAGTTTGTCAATGTAGTTCAGATGTCAAAACCGCTTCCGGATGC 1680  
QY 1681 CGCTGCAAGCAAGTGTCAACCAAGAGTCCGCTGCTACTGCTGTCCGAGAGTGT 1740  
DB 1681 CGCTGCAAGCAAGTGTCAACCAAGAGTCCGCTGCTACTGCTGTCCGAGAGTGT 1740  
QY 1741 GACCCGTGACCTGTCTTACTGTGTGAGCGCGTGCACATTTGGAGACAGTAAATGTGTC 1800  
DB 1741 GACCCGTGACCTGTCTTACTGTGTGAGCGCGTGCACATTTGGAGACAGTAAATGTGTC 1800  
QY 1801 TGCAAGAACTGCAAGTATTCAGCGGGGCTCAAAAGACATCTATGCTGACCATCTGAC 1860  
DB 1801 TGCAAGAACTGCAAGTATTCAGCGGGGCTCAAAAGACATCTATGCTGACCATCTGAC 1860  
QY 1861 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAAAATGAAATTCATCTCAGAA 1920  
DB 1861 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAAAATGAAATTCATCTCAGAA 1920  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGTGAAGTGAACAGAAAGGAAATGTATGATAA 1980  
DB 1921 TACTGTGAGAGATTAATTTCTCAAGTGAAGTGAACAGAAAGGAAATGTATGATAA 1980  
QY 1981 TACATGTGACGCTTCTGTTCACCTGAACATGATTTGTGTGTGATCAACCCGCAAG 2040  
DB 1981 TACATGTGACGCTTCTGTTCACCTGAACATGATTTGTGTGTGATCAACCCGCAAG 2040  
QY 2041 GGTAAACAAAATTCGTTTGCAATCAATTCGCTGCTATGCAAAAGTTATG 2100  
DB 2041 GGTAAACAAAATTCGTTTGCAATCAATTCGCTGCTATGCAAAAGTTATG 2100  
QY 2101 AATGCTTAAGCGTATCAAGATAGTATTTTGGCAAGAGCCATCCAGCTGCCGA 2160  
DB 2101 AATGCTTAAGCGTATCAAGATAGTATTTTGGCAAGAGCCATCCAGCTGCCGA 2160  
QY 2161 GAGCTGTATTTTGAATTAAGATACAGCAGGCTGATGCTGAGATGTCCGATCGAA 2220  
DB 2161 GAGCTGTATTTTGAATTAAGATACAGCAGGCTGATGCTGAGATGTCCGATCGAA 2220

QY 2221 AGAGAAATGGAATCCCTTGA 2241  
DB 2221 AGAGAAATGGAATCCCTTGA 2241  
RESULT 3  
AY399443  
LOCUS 2241 bp DNA linear GSS 15-DEC-2003  
DEFINITION Mus musculus E2H2 gene, VIRTUAL TRANSCRIPT, partial sequence,  
AY399443  
ACCESSION AY399443  
VERSION AY399443.1 GI:39755432  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2241)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
Todd, M.A., Tanenbaum, D.M., Clivello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
TITLE 2 (bases 1 to 2241)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
Todd, M.A., Tanenbaum, D.M., Clivello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.  
Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source location/Qualifiers  
1..2241  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..>2241  
/gene="E2H2"  
/locus\_tag="HMC0225"  
ORIGIN  
Query Match 73.8%; Score 1654.8; DB 14; Length 2241;  
Best Local Similarity 77.4%; Pred. No. 0; Mismatches 507; Indels 0; Gaps 0;  
Matches 1734; Conservative 0  
QY 1 ATGGCCAGACTGGAGAAATCTGAGAAAGGACGATTTGTGCGGAAAGCTGTAAAA 60  
DB 1 ATGGCCAGACTGGAGAAATCTGAGAAAGGACGATTTGTGCGGAAAGCTGTAAAA 60  
QY 61 TCAAGTACATGCACTGACAGCACTCAAGAGTTTCAAGCAAGCTGATGAAAGTAAAGT 120  
DB 61 TCAAGTACATGCACTGACAGCACTCAAGAGTTTCAAGCAAGCTGATGAAAGTAAAGT 120  
QY 121 ATGTTTACGTCACATCGTCAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATGG 180  
DB 121 ATGTTTACGTCACATCGTCAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATGG 180  
QY 181 AAACAGCAAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGCGCGGACT 240  
DB 181 AAACAGCAAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGCGCGGACT 240  
QY 241 AGGAGTGTGGTGAACAAGTGAAGCTTGTGCAACCAAGTATCCCATTTAAAGCT 300  
DB 241 AGGAGTGTGGTGAACAAGTGAAGCTTGTGCAACCAAGTATCCCATTTAAAGCT 300  
QY 301 CTGAATGAGTTCCTTCAAGTACCAATTAATGATTTCTGTGTCTCCCTACAGCAATTTT 360

[illegible]

Db	1381	AAAAAAGTAGACAGAGGTATGAGTTTGAAGTCAAGAGTCCAGTATCATAGCACTGTT	1440
Qy	1441	CCCGCTGAGGATGTGGATTACTCTCCCAAGAAAAAGAGGAAACCCGGTGTGGGCT	1500
Db	1441	CCCACTGAGAGTAGTAGCACTCTCCCAAGAAAGAAAGAAACATCGGTGTGGGCT	1500
Qy	1501	GCAACACTGAGAAAGATACAGCTGAAAAAGAGCGGCTCTTAACCATGTTTCAACTAT	1560
Db	1501	GCACACTGCGAAGAAAGATACAACTGAAAAAGAGCGGCTCTCTAACCATGTTTCAACTAT	1560
Qy	1561	CAACCCCTGATCATCCAGCGGACGCTTGTGACGTTTGTGCCCTTGTGTATGCAAA	1620
Db	1561	CAACCCCTGTGACCATCCAGCGGACGCTTGTGACGTTTGTGCCCTTGTGTATGCAAA	1620
Qy	1621	AATTTTGTGAAAAAGTTTGTGCATGTAGTTCAGAGTCAAAACCGCTTTCGGAATGT	1680
Db	1621	AATTTTGTGAAAAAGTTTGTGCATGTAGTTCAGAGTCAAAACCGCTTTCGGAATGT	1680
Qy	1681	CGCTGCAAAAGCAGTGCACAACCAAGACGTGCCGTGTACTCGGCTGTCCGAGTGT	1740
Db	1681	CGGTGCAAAAGCAGATGCCAACCAAGATGTCAATGTCTACTCGGCTGTCCGAGATGT	1740
Qy	1741	GACCTTGAACCTCTGTCTTACTTGTGGAGCGGCTGACCAATTGGGACATTAATAAATGTGTCC	1800
Db	1741	GACCTTGAACCTCTGTCTTCAAGTGTGAGCTGTGACCAATTGGGACATTAATAAATGTATCC	1800
Qy	1801	TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAAGCATCTATTGCTGGCACCATCTGAC	1860
Db	1801	TGTAAAGACTGTAGCATTCACGCGGGGCTTAATAAAGCATTTACTGTGGCACCGCTGAT	1860
Qy	1861	GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAGAAAAATGAATTCTTCCGAA	1920
Db	1861	GTGGCAGGCTGGGGGCACTTTATCAAAAGATCTGTGCAGAAAAATGAATTCTTCCGAA	1920
Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACGAAGAGGGAAGTATGATATA	1980
Db	1921	TACTGTGGGAGATTAATTTCTCAAGATGAAGCAGACAGAGAGGAAGTATGACAAA	1980
Qy	1981	TACATGTGCAGCTTCTGTTCAACTTGAACAATGATTTGTGTGATGCAACCCGCAAG	2040
Db	1981	TACATGTGCAGCTTCTGTTCAACTTGAACAATGATTTGTGTGATGCAACCCGCAAG	2040
Qy	2041	GGTACCAAAATTCGTTTTCGCAATCATTCGTTAAATCCAACTGTATGCAAAAGTTATG	2100
Db	2041	NN	2100
Qy	2101	ATGTGTAAACGTGATCAACAGATAGGTATTTTCCAAAGAGGCATTCAGACTGCGCAA	2160
Db	2101	NN	2160
Qy	2161	GAGCTGTTTTTGTGATTAACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGCATCGAA	2220
Db	2161	NN	2220
Qy	2221	AGAGAAATGGAATCCCTTGA 2241	
Db	2221	CGAGAAATGGAATCCCTTGA 2241	

RESULT 4	AY399442	2241 bp	DNA	linear	GSS 15-DEC-2003
LOCUS	AY399442				
DEFINITION	Pan troglodytes E2H2 gene, VIRUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY399442				
VERSION	AY399442.1				
KEYWORDS	GI:39755431				
SOURCE	GSS.				
ORGANISM	Pan troglodytes (chimpanzee)				
	Pan troglodytes				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.				
REFERENCE	1 (bases 1 to 2241)				

AUTHORS Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeilio,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Srinisky,J.D.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 2241)  
AUTHORS Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeilio,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Srinisky,J.D.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source location/Qualifiers  
1..2241  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..>2241  
/gene="RZM2"  
/locus\_tag="HCM0225"  
ORIGIN  
Query Match 71.0%; Score 1590.2; DB 14; Length 2241;  
Best Local Similarity 71.0%; Pred. No. 0;  
Matches 1592; Conservative 0; Mismatches 649; Indels 0; Gaps 0;  
QY 1 ATGGCCAGACTGGGAGAAATCTGAGAGGACCACTTTGGCGGAGCGTGTAAA 60  
DB 1 ATGGCCAGACTGGGAGAAATCTGAGAGGACCACTTTGGCGGAGCGTGTAAA 60  
QY 61 TCAGAGTCAATGCGACTGAGACGCTCAAGAGTTCAAGCGAGCTGATGAATGAAGT 120  
DB 61 TCAGAGTCAATGCGACTGAGACGCTCAAGAGTTCAAGCGAGCTGATGAATGAAGT 120  
QY 121 ATGTTAGTTCATTCGTGCAAAATTTTGAAGAAGACGAATCTTAAACCAAGATGG 180  
DB 121 NNN 180  
QY 181 AAACAGCGAAGATACAGCTGTGCAATCTGACTCTGTGAGCTCATTTGCGCGGACT 240  
DB 181 NNN 240  
QY 241 AGGAGTGTTCGCTGACGAGTGAATTTTCCAAACAAGTCAATCCATTTAAAGCT 300  
DB 241 NNN 300  
QY 301 CTGAATGCACTGCTTCACTACCACTAATGTAATCTTGGTCTCCCTACAGCAGATTTT 360  
DB 301 NNN 360  
QY 361 ATGTGGAAGATGAACCTGTTTACATACTTCTTATATGAGAGATGAAGTTTGAAT 420  
DB 361 NNNNGTGAAGATGAACCTGTTTACATACTTCTTATATGAGAGATGAAGTTTGAAC 420  
QY 421 CAGGATGTAATTTGAAGAACTAAATAAATTTATGATGGAAGTAAACCGGAT 480  
DB 421 CAGGATGTAATTTGAAGAACTAAATAAATTTATGATGGAAGTAAACCGGAT 480  
QY 481 AGAGATGTGGTTTAAATGATGAATTTTGTGAGAGTGTGATCCCTGGTCA 540  
DB 481 AGAGATGTGGTTTAAATGATGAATTTTGTGAGAGTGTGATCCCTGGTCA 540  
QY 541 TATTAATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 541 TATTAATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 601 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGACCGCCCACTCGGAAATTTCTTCT 660

DB 601 NNN 660  
QY 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCAATTAAGGCGACAGCAGAACTA 720  
DB 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCAATTAAGGCGACAGCAGAACTA 720  
QY 721 AAAGAAATTAATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGAATGT 780  
DB 721 AAAGAAATTAATTAAGAACTCAACGAGAGCTCCAGGCGCACTTCTCTGAATGT 780  
QY 781 ACCCCCAATAGATGACCAAAATGCTAAATCTGTTCAAGAGAGCAAGCTTACATCC 840  
DB 781 ACCCCCAATAGATGACCAAAATGCTAAATCTGTTCAAGAGAGCAAGCTTANNNTCC 840  
QY 841 TTTCAATGCTTTTCTGATGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 900  
DB 841 TTTCAATGCTTTTNN 900  
QY 901 GCAACACCAACACTTATTAAGCGAAGAACACAGAAACAGCTTGAACAACAACTTGT 960  
DB 901 NNN 960  
QY 961 GGAACACAGTGTACAGCAATTTGAGAGGAGCAAAAGAGTTGCTGCTCTCACCGCT 1020  
DB 961 NNN 1020  
QY 1021 GAGCGATTAAGACCCCAACAAACGTCAGAGGCGCGAGAGAGACGCTTCCCAAT 1080  
DB 1021 GAGCGATTAAGACCCCAACAAACGTCAGAGGCGCGAGAGAGACGCTTCCCAAT 1080  
QY 1081 AAAGATGACAGGCGCGAGACCCCAACCAATATGTCGTAATCAAGATGACAGCT 1140  
DB 1081 AAAGATGACAGGCGCGAGACCCCAACCAATATGTCGTAATCAAGATGACAGCT 1140  
QY 1141 GATAGGAGAGAGGAGCTGAACCGGGGAGAGAAATGATTAAGAAAGAGAGAG 1200  
DB 1141 GATAGGAGAGAGGAGCTGAACCGGGGAGAGAAATGATTAAGAAAGAGAGAG 1200  
QY 1201 AAAGATGAACCTTGAAGCTCTCTGAGCAAAATCTCGGTGCAAAACCAATTAAGATG 1260  
DB 1201 AAAGATGAACCTTGAAGCTCTCTGAGCAAAATCTCGGTGCAAAACCAATTAAGATG 1260  
QY 1261 AAGCAATATTTGAACCTCTCTGAGAAATGTCGATGTCGTCGTCGTCGTCGTCGTCGTC 1320  
DB 1261 AAGCAATATTTGAACCTCTCTGAGAAATGTCGATGTCGTCGTCGTCGTCGTCGTCGTC 1320  
QY 1321 AAGTCTCATTTGCACTTATGACAAATTTCTGTGCAATTTGCTAGGTTAATTTGGAGC 1380  
DB 1321 AAGTCTCATTTGCACTTATGACAAATTTCTGTGCAATTTGCTAGGTTAATTTGGAGC 1380  
QY 1381 AAAACATGTAGACAGTGTATGATGATTAAGTCAAGATCTAGCATAGCTTCACT 1440  
DB 1381 AAAACATGTAGACAGTGTATGATGATTAAGTCAAGATCTAGCATAGCTTCACT 1440  
QY 1441 CCGGCTGAGATGTCGATCTCTCTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB 1441 CCGGCTGAGATGTCGATCTCTCTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 GCAACCTGAGAGATGACGTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
DB 1501 GCAACCTGAGAGATGACGTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
QY 1561 CAACCTGTATCATCCAGGAGAGCTTGTGACAGTTGTCGTCGTCGTCGTCGTCGTCGTC 1620  
DB 1561 CAACCTGTATCATCCAGGAGAGCTTGTGACAGTTGTCGTCGTCGTCGTCGTCGTCGTC 1620  
QY 1621 AATTTTGTGAAAGTTTGTCAATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1680  
DB 1621 AATTTTGTGAAAGTTTGTCAATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1680  
QY 1681 CGCTGCAAGAGCAGTGCAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740



Db	1681	CGCTGCAAAAGCACAGTGCACCAACCAAGAGTGCCTGCTACCTGCTGCTGCAGAGAGTGT	1740
Qy	1741	GACCCCTGACCTCTGTCTTACTTTGTGGAAGCCGTGCACCATTTGGGACAGTAAAAAATGTGTCC	1800
Db	1741	GACCCCTGACCTCTGTCTTACTTTGTGGAAGCCGTGCACCATTTGGGACAGTAAAAAATGTGTCC	1800
Qy	1801	TGCAAGAACTGCAAGTATCAGCGGGGGCTCCAAAAAGCATCTATTGCTGTGGACCACTTGAC	1866
Db	1801	TGCAAGAACTGCAAGTATCAGCGGGGGCTCCAAAAAGCATCTATTGCTGTGGACCACTTGAC	1866
Qy	1861	GTGGCAGGCTGGGGATTTTATCAAAAGTCCTGTGCAGAAAAATGAATTATCTCAGAA	1920
Db	1861	GTGGCAGGCTGGGGATTTTATCAAAAGTCCTGTGCAGAAAAATGAATTATCTCAGAA	1920
Qy	1921	TACTGTGAGAGATTATTTCTCAGAGTAAAGCTGCAGAGAGGGAAAGTATGATTAA	1980
Db	1921	TACTGTGAGAGATTATTTCTCAGAGTAAAGCTGCAGAGAGGGAAAGTATGATTAA	1980
Qy	1981	TACTGTGCACTTTCCTTCAACTTGAACAATGATTTTGTGTGAGATGCAACCCGCAAG	2040
Db	1981	TACTGTGCACTTTCCTTCAACTTGAACAATGATTTTGTGTGAGATGCAACCCGCAAG	2040
Qy	2041	GCTAACAAAATCTGTTTGCAAAATCATTCGTGTAATCCAACTGCTATGCAAAAGTTATG	2100
Db	2041	NN	2100
Qy	2101	ATGCTTAAACGGTGATCAGAGATAGATATTTTGTCCAGAGAGCCATCAGACTGGCGAA	2166
Db	2101	NN	2166
Qy	2161	GAGCTGTTTTTGAATTACAGATACAGCCAGGCTGATGCCCTGAAAGTATGTGGCATCGAA	2220
Db	2161	NN	2220
Qy	2221	AGAGAAATGAAATCCCTTGA	2241
Db	2221	AGAGAAATGAAATCCCTTGA	2241

RESULT 5	
AK153835	1238 bp mRNA linear HTC 21-SEP-2005
LOCUS	Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
DEFINITION	full-length enriched library, clone:c920016A05 product:enhancer of
	zeete homolog 2 (Drosophila), full insert sequence.
ACCESSION	AK153835
VERSION	AK153835.1 GI:74150295
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
	Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	
AUTHORS	1 Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	10349636
REFERENCE	
AUTHORS	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komo,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
TITLE	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	11042159
REFERENCE	
AUTHORS	3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Komo,H., Akiyama,J., Nishi,K., Kitsuma,T., Tashiro,H., Itoh,M., Sunji,N., Ishii,Y., Nakamura,S., Hazana,M., Nishino,T., Harada,A., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuno,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.

TITLE	JOURNAL	PUBLISHED	REFERENCE
RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer	Genome Res.	10 (11), 1757-1771	(2000)
11076861			
4			
Kawai, T., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kodora, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudil, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldirelli, R., Barsch, G., Blake, J., Botfield, D., Bojunga, N., Carninci, P., de Bonaldo, M. P., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hall, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Schonwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohenaki, S. and Hayashizaki, Y.			
RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium	Functional annotation of a full-length mouse cDNA collection	Nature	409 (6821), 685-690 (2001)
11217851			
5			
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldirelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanpin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusc, V., Chothia, C., Cobani, L. B., Cousins, S., Dalla, E., Dragan, T. A., Fletcher, C. P., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kawai, A., Kawaji, H., Kawasawa, Y., Kedzeraki, R. M., King, B. L., Konagawa, A., Kurochkin, I. V., Lee, Y., Lenhard, P. A., Lyons, P. A., Maglott, D. R., Maltis, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Seto, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakakawa, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.			
PANTOM Consortium	Analysis of the mouse transcriptome based on functional annotation	Nature	420 (6915), 563-573 (2002)
12466851			
6			
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Kavasi, T., Lenhard, B., Wells, C., Kodzumi, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aldred, R., Allen, J. E., Ambesi-Impombato, A., Apweiler, R., Attalaya, R. N., Bailey, T. L., Banasi, K. P., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Fukui, S., Gariboldi, M.,			



Georgii-Hemming, P., Gingers, T.R., Gobjoori, T., Green, R.E.,  
Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N.,  
Hill, D., Humnietek, L., Iacono, M., Ileo, K., Iwama, A., Ishikawa, T.,  
Jait, M., Kanapin, A., Katoh, M., Kawasawa, Y., Keisio, J., Kitamura, H.,  
Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K.,  
Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J.,  
Liu, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L.,  
Masuda, H., Matsuzawa, S., Miki, H., Mlynore, P., Miyake, S.,  
Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H.,  
Ng, P., Nilsson, R., Niehughi, S., Nishikawa, S., Nori, P., Ohara, O.,  
Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavani, G.,  
Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z.,  
Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandler, A.,  
Schneider, C., Schonbach, C., Seiliguchi, K., Semple, C.A., Seno, S.,  
Seesa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D.,  
Sinclair, B., Sperling, S., Stupka, B., Suglura, K., Sultana, R.,  
Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S.,  
Tegner, J., Teichmann, S.A., Ueda, H., R. van Nimwegen, B., Verardo, R.,  
Wei, C.L., Yagi, K., Yamaniishi, H., Zaborovky, B., Zhu, S., Zimmer, A.,  
Hilde, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T.,  
Brice, V., Quackenbush, J., Wahlested, C., Mattick, J.S., Hume, D.A.,  
Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, K., Itoh, M.,  
Suzuki, M., Aoki, J., Aikawa, T., Iida, J., Imanura, K., Itoh, M.,  
Kato, T., Kawaji, H., Kawagashita, N., Kawashima, T., Kojima, M.,  
Kondo, S., Komori, H., Nakano, K., Niimoriya, N., Nishio, T., Okada, M.,  
Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K.,  
Watahiki, A., Okamura-Ono, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE

The transcritom landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)  
16141072

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M.,  
7

CONSRNM	RIKEN Genome Exploration Research Group
TITLE	Antisense transcription in the mammalian transcriptome
JOURNAL	Science 309 (5740), 1564-1566 (2005)
PUBMED	16141073
REFERENCE	8 (bases 1 to 1238)
AUTHORS	Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K.,

REFERENCE  
AUTHORS

8 (pages 1 to 1238)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K.,  
Hori, F., Iida, Y., Imanura, K., Imocani, K., Itoh, M., Kanagawa, S.,  
Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Nishimoto, N.,  
Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D.,  
Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Wataniki, A.,  
Muramatsu, M., and Hayashizaki, Y.

QY 721 AAGGAAAAATATAAGAACTCAACGAGAGAGCTCCAGGCGCACTTCTCTGAATGT 780  
DB 786 AAGGAAAAATATAAGAACTCAACGAGAGAGCTCCAGGCGCTCTGCTCTGAATGT 845  
QY 781 ACCCCCAACATAGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 840  
DB 846 ACTCAAAACATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 905  
QY 841 TTTCAATAGCTTTCTGTAAGGCGAATGTTTAAATATGATGCTGCTTCAATCTTTCAT 900  
DB 906 TTTCAATAGCTTTCTGTAAGGCGAATGTTTAAATATGATGCTGCTTCAATCTTTCAT 965  
QY 901 GCAACACCCCAACATTAATAGCGAGAAACAGAAACAGCTCTAGACAAACCTTGT 960  
DB 966 GCAACACCCCAACATTAATAGCGAGAAACAGAAACAGCTTGTGAACAAGCTTGT 1025  
QY 961 GGAACCAAGTGTATACAGCAATTTGAGGAGAGCAAAAGATTTGCTGCTCTCAACGCT 1020  
DB 1026 GGAACCAAGTGTATACAGCAATTTGAGGAGAGCTGTAAGGAGTGTGCTGCTCTCAACGCT 1085  
QY 1021 GAGGCGATTAAGAACCCCAACCAAAAGTCCAGAGGCGCAGAGAGAGAGCGCTTCCAT 1080  
DB 1086 GAGGCGATTAAGAACCAACCTTAAGCGCCAGGCGCGCAGAGAGAGAGAGCTTCCAT 1145  
QY 1081 AACAGTACAGAGCCCAACCCCAACCTTAATATGCTGTAATCAAGATACAGACAGT 1140  
DB 1146 AACAGTACAGAGCCCAACCCCAACCTTAATATGCTGTAATCAAGATACAGACAGT 1205  
QY 1141 GATAGGAG 1173  
DB 1206 GACAG 1238

RESULT 6  
CR990276 873 bp mRNA linear EST 28-JUN-2005

LOCUS CR990276 RZPD no. 9016 Homo sapiens cDNA clone RZPD9016B0143 5',  
DEFINITION mRNA sequence.

ACCESSION CR990276  
CR990276.1 GI:68284161

VERSION CR990276.1 GI:68284161  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo;  
1 (bases 1 to 873)

REFERENCE Heil, O., Ebert, T., Hennig, S., Henze, S., Radelof, U., Schneider, D.  
AUTHORS and Korn, B.  
TITLE Human T-Lymphocytes library  
JOURNAL Unpublished (2005)  
COMMENT Contact: Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Email: www.rzpd.de  
RZPD: RZPD9016B0143.  
RZPDLIB: (Human T-Lymphocytes) RZPD LIB No. 9016  
http://www.rzpd.de/cgi-bin/products/set.cgi?libno=9016 Contact:  
Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 100  
Fax: +49 30 32639 111  
www.rzpd.de

FEATURES  
source  
1. .873  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/clone="RZPD9016B0143"  
/issue\_type="T-Lymphocytes"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="RZPD no. 9016"  
/note="Vector: pOB80LSN\_cloned; Site\_1: SalI; Site\_2:  
NotI; vector:  
http://www.rzpd.de/info/vectors/pOB80LSN\_cloned\_pic.shtml  
; 1st strand cDNA was prepared from mRNA obtained from  
human T-Lymphocytes with a NotI - oligo(dT) primer [5'  
GACTAGTTCAGATTCGGAGCGCGCCCTTTTCTTTTCTTTT 3'].  
Double-stranded cDNA was ligated into the NotI and SalI sites  
of the pOB80LSN\_cloned vector"

Query Match 38.9%; Score 871.4; DB 8; Length 873;  
Best Local Similarity 99.9%; Pred. No. 5e-223;  
Matches 872; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 712 GAAGAACTAAAGAAAAATATAAGAACTCAACGAGAGAGCTCCAGGCGCACTTCTCT 771  
DB 1 GAAGAACTAAAGAAAAATATAAGAACTCAACGAGAGAGCTCCAGGCGCACTTCTCT 60  
QY 772 CTGAAATGTACCCCAACATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGC 831  
DB 61 CTGAAATGTACCCCAACATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGC 120  
QY 832 TTACAGCTCTTCAATAGAGCTTTCTGTAAGGAGATGTTTAAATATGATGCTGCTCTCAT 891  
DB 121 TTACAGCTCTTCAATAGAGCTTTCTGTAAGGAGATGTTTAAATATGATGCTGCTCTCAT 180  
QY 892 CTTTTCATGAGCAACCCCAACATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
DB 181 CTTTTCATGAGCAACCCCAACATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 952 AAACCTTGTGAGCAACAGTGTATACAGCAATTTGAGAGAGAGAGAGAGAGAGAGAG 1011  
DB 241 AAACCTTGTGAGCAACAGTGTATACAGCAATTTGAGAGAGAGAGAGAGAGAGAGAG 300  
QY 1012 CTCAAGCTGAGGAGATTAAGAACCCCAACCAAAAGTCCAGAGGCGCAGAGAGAGAGAG 1071  
DB 301 CTCAAGCTGAGGAGATTAAGAACCCCAACCAAAAGTCCAGAGGCGCAGAGAGAGAGAG 360  
QY 1072 CTTCCCAATACAGTACAGAGCCCAACCCCAACCTTAATATGCTGTAATCAAGAGAT 1131  
DB 361 CTTCCCAATACAGTACAGAGCCCAACCCCAACCTTAATATGCTGTAATCAAGAGAT 420  
QY 1132 ACAGACAGTGTAG 1191  
DB 421 ACAGACAGTGTAG 480  
QY 1192 GAAG 1251  
DB 481 GAAG 540  
QY 1252 ATAAAGATGAG 1311  
DB 541 ATAAAGATGAG 600  
QY 1312 TCAATGTTTAAAGTCTCATATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371  
DB 601 TCAATGTTTAAAGTCTCATATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 1372 ATTGGAGCAAAACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431  
DB 661 ATTGGAGCAAAACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 1432 GCTCCAGCTCCCGCTGAG 1491  
DB 721 GCTCCAGCTCCCGCTGAG 780



cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCM2438 row: 9 column: 10  
High quality sequence stop: 677.  
Location/Qualifiers

FEATURES  
source

```
1..956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6267681"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 109"
/notes="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
```

## ORIGIN

Query Match 37.9%; Score 848.4; DB 3; Length 956;  
Best Local Similarity 98.1%; Pred. No. 8.2e-217;  
Matches 869; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```
361 ATGTGTGAAGATGAAGAACTGTTTACATTAACATTCCTTATATGAGATGAAGTTTATAGAT 420
Db 72 AAGGTGGAAGATGAAGAACTGTTTACATTAACATTCCTTATATGAGATGAAGTTTATAGAT 131
Qy 421 CAGATGTGATCTTCATGATGAAGAACTATATAAAATTTGATGGGAAAGTACAGGGGAT 480
Db 132 CAGATGTGATCTTCATGATGAAGAACTATATAAAATTTGATGGGAAAGTACAGGGGAT 191
Qy 481 AGAGATGTGGGTTTATATAATGATGAAGAAATTTTGGAGTGGTGAATGCCCTTGCTCAA 540
Db 192 AGAGATGTGGGTTTATATAATGATGAAGAAATTTTGGAGTGGTGAATGCCCTTGCTCAA 251
Qy 541 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 252 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
Qy 601 AAAGATCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 312 AAAGATCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
Qy 661 GATTAATATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGGGACAGAGAAAGACTA 720
Db 372 GATTAATATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGGGACAGAGAAAGACTA 431
Qy 721 AAGGAAAAATATAAAGAACTACCGAAGAGAGCTCCAGGGGCACTTCTCTGGAATGT 780
Db 432 AAGGAAAAATATAAAGAACTACCGAAGAGAGCTCCAGGGGCACTTCTCTGGAATGT 491
Qy 781 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 492 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 551
Qy 841 TTTCATAGCTTTTCTGAGGCGATGTTTAAATATGATGATGATGATGATGATGATGATGATGAT 900
Db 552 TTTCATAGCTTTTCTGAGGCGATGTTTAAATATGATGATGATGATGATGATGATGATGATGAT 611
Qy 901 GCAACACCAACACTTATAGCGGAAGAACAGAAAGAGCTTATAGACAAACCTTGT 960
Db 612 GCAACACCAACACTTATAGCGGAAGAACAGAAAGAGCTTATAGACAAACCTTGT 671
Qy 961 GACCAACGTGTATACCAAGATTTGAGGAGAGAAAGAGTTTGTCTGCTCTCACCGCT 1020
```

```
Db 672 GACACACAGTGTATCCAGCATTTTGGAGGAGCAAAAGAGTTTGTCTGCTCTCACCGCT 731
Qy 1021 GACCGGATTAAGACCCCAACCAAAAGCTCCAGAGGCGCGAGAAAGAGCGCTTCCCAAT 1080
Db 732 GACGGATTAAGACCCCAACCAAAAGCTCCAGAGGCGCGAGAAAGAGCGCTTCCCAAT 791
Qy 1081 AACAGTATGACGCGCCAGACACCCCAACATTAATATGATGATGATGATGATGATGATGATGAT 1140
Db 792 AACAGTATGACGCGCCAGACACCCCAACATTAATATGATGATGATGATGATGATGATGATGAT 851
Qy 1141 GATAGGAGACAGGAGACTGAACAGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1200
Db 852 GATAGGAGACAGGAGACTGAACAGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 910
Qy 1201 AAAGATGAACCTTCAGACTCTCTGAGAACAAATTTCTGGTGTCAA 1246
Db 911 AAAGATGAACCTTCAGACTCTCTGAGAACAAATTTCTGGTGTCAA 956
```

RESULT 9  
BU556770 972 bp mRNA linear EST 16-SEP-2002  
AGENCOURT 10186850 NIH\_MGC 109 Homo sapiens cDNA clone  
IMAGE:6584530 5', mRNA sequence.  
ACCESSION BU556770.1 GI:22907066  
VERSION BU556770  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 972)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCM2795 row: a column: 10  
High quality sequence start: 18  
High quality sequence stop: 662.  
Location/Qualifiers

FEATURES  
source

```
1..972
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6584530"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 109"
/notes="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

## ORIGIN

Query Match 37.8%; Score 846.2; DB 3; Length 972;  
Best Local Similarity 96.5%; Pred. No. 3.2e-216;  
Matches 874; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

```
580 GAAGAAAGAAAGAAAGCAAGATCTGAGAGTACCGAGATGATTAAGAAAGCGCG 639
```



Db	241	TAATAAAATTGATGAGGAAAGTACACGGGGATGAGAAATGCGGGTTATTAATAATGATGAA	300
Qy	509	TTTTTGTGGAAGTTGCTGAATGCGCTTGGTCAATATATGATGATGACGATGATGATG	568
Db	301	TTTTTGTGGAAGTTGCTGAATGCGCTTGGTCAATATATGATGATGACGATGATGATG	360
Qy	569	GAGACGATCTCGAAGAAAGAGAAAGAAAGCAAGAAAGATCTGAGAGATCACCGAGTGA	628
Db	361	GAGACGATCTCGAAGAAAGAGAAAGAAAGCAAGAAAGATCTGAGAGATCACCGAGTGA	420
Qy	629	AAGAAGCCGCCCACTTCGGAATTTCTTCGTATTAATTTTGGACCAATTTCTTCA	688
Db	421	AAGAAGCCGCCCACTTCGGAATTTCTTCGTATTAATTTTGGACCAATTTCTTCA	480
Qy	689	TGTTTCCAGATTAAGGGCAACAGCAAGAACTTAAGGAAAAATATTAAGAACTCACCGAAC	748
Db	481	TGTTTCCAGATTAAGGGCAACAGCAAGAACTTAAGGAAAAATATTAAGAACTCACCGAAC	540
Qy	749	AGCAGCTCCCGAGGGCACTTCTCCTGAAATGTAACCCCAATGATGATGACCAATGCTA	808
Db	541	AGCAGCTCCCGAGGGCACTTCTCCTGAAATGTAACCCCAATGATGACCAATGCTA	600
Qy	809	AATCTGTTCAGAGAGAGCAAAAGCTTACACTCTTTTCATACGCTTTCTGTAGGCGATGT	868
Db	601	AATCTGTTCAGAGAGAGCAAAAGCTTACACTCTTTTCATACGCTTTCTGTAGGCGATGT	660
Qy	869	TTAAATATGACTGCTTCTCTTACATCTTTTCATGCAACCCCAACTTATTAAGCGGANA	928
Db	661	TTAAATATGACTGCTTCTCTTACATCTTTTCATGCAACCCCAACTTATTAAGCGGANA	720
Qy	929	ACACAGAAACAGCTCTTACAAACAAACCTTGAGCAACAGTTTACAGAGATTTGGAG	988
Db	721	ACACAGAAACAGCTCTTACAAACAAACCTTGAGCAACAGTTTACAGAGATTTGGAG	780
Qy	989	GAGCAAGAGATTTGCTGCTGCTCTCAACGCTGAGCGAGTAAAGACCCCAACAAACCTC	1048
Db	781	GAGCAAGAGATTTGCTGCTGCTCTCAACGCTGAGCGAGTAAAGACCCCAACAAACCTC	840
Qy	1049	CA 1050	
Db	841	CA 842	
RESULT 11			
CR857825		4643 bp mRNA linear HTC 12-NOV-2004	
LOCUS	Pongo pygmaeus mRNA; cDNA DKFp469N1815 (from clone DKFp469N1815)		
DEFINITION	CR857825		
ACCESSION	CR857825.1	GI:55726634	
VERSION			
KEYWORDS	HTC.		
SOURCE	Pongo pygmaeus (orangutan)		
ORGANISM	Pongo pygmaeus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Pongo.		
AUTHORS	1 (bases 1 to 4643)		
CONSRM	Kocher, K., Beyer, A., Mewes, H.W., Weill, B., Amlid, C., Oeanger, A., Fobo, G., Han, M. and Wiemann, S.		
JOURNAL	The German cDNA Consortium		
COMMENT	Direct Submission		
	Submitted (12-NOV-2004) MIRS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY		
	Cloned by S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMRZ (Biomedical Research Center at the Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFp469N1815) is available at the RZPD Deutsches Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:		
	http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneid=DKFp469N1815		
	Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.		

FEATURES

SOURCE

1. .4643 location/Qualifiers

/organism="Pongo pygmaeus"

/mol\_type="RNA"

/db\_xref="taxon:9600"

/clone="DKFZp469N1815"

/issue\_type="Kidney"

/clone\_lib="469 (synonym: pKid). Vector pSport1\_5fi; host DH10B; sites SfiIa + SfiIb"

/dev\_stage="adult"

/note="enhancer of zeste homolog 1 (Homo sapiens)"

1. .4643

/gene="DKFZp469N1815"

119. 2362

/gene="DKFZp469N1815"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAH90081.1"

/db\_xref="GI:55726635"

/translation="MEIINPPTSKCTIKYMKRKVSEYWRRLQLRILOANMGAKLYLVNFAVKQKTOIINEBKKLAVQVPSKRPVCGHPIKKCTESIIPGASQHMRSISNTVALVPIWTSWSPLOQNFWEDETVLCNIPYMDYKEDETIEELINNYDGVKHEBEINPGSVLISDAVFLFLVDALNQPYDDEEGENDVSDGQDSKEDLPYTRKRRAAIBENKSSKKQFPNDMIPSAIASMPENGVPMDKRVERLITMSDPNALPQCTQPCCYLAIRECDPLCTCGASEHMDCKVSCNKCSIORGLKHLIAPSDVAWGTEIKSEYVQKNEFISRYCGELISQDSADRGRKYDKMSSFLFNINLMDPVDAITRKMKIPRAHSVNPNCYATVAVWVNGDHRIGIFAKRAIQAGEELFLYRYSQADALKVGGEBREIDVL"

ORIGIN

Query Match 37.1%; Score 832.4; DB 6; Length 4643;

Best Local Similarity 64.0%; Pred. No. 2.7e-212;

Matches 1423; Conservative 0; Mismatches 726; Indels 75; Gaps 8

Db 55 GTAAATCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAACGAGCTGATGAAGTA 114

173 GTGAATCTGTAATCATGCGACTTCGACAACTTAAACGGCTTCAAGCAAAATATGGTGCA 232

Qy 115 AAGGATATGTTAGTTCCATATGTCAGAAAATTTTGAAAGAAGCGAATCTTAAACCA 174

Db 233 AAGCTTTGTAATGTCGCAAAATTTTGCAAAAGTTCAAGAAAAACCAATCTCTCAATGA 292

Qy 175 GAATGGAAGACGCGAAGGATACAGCTGTGACATCTCGACTTGTGAGCTCATTTGCGC 234

Db 293 GAATGGAAGAGCTTCTGTCTCAACCTTTCAGTCATGAAAGCCTGTGTGTGGAACCTT 352

Qy 235 GGGACTAGGAGTGTTC--GATGACAGTGACTTGAAATTTTCCAACAACAAGTCATCCCA 291

Db 353 TTTCTCAAAAAGTGAACATAGAGGAGATTTTCCGGGATTTGCAAGCAACATATGTTA 412

Qy 292 TTAAAGACTCGAATGACAGTTGCTTCAGTACCATAATGTAATCTTGGTCTCCCTCAACG 351

Db 413 ATGAGGTCACTGGAACACAGTGTGATGTGTTCCCATCAAGTATCTCTGTCTCCCA 472

Qy 352 CAGAATTTTATGATGGAAGATGAACCTGTTTACATACATTCCTTATATGAGAGATGAA 411

Db 473 CAGAACTTATAGGTGAAGATGAGACGGTTTTTGTGCAATATTCCTTACATGGAGATGAA 532

Qy 412 GTTTTATGATCAGATGTGACTTTCATTTGAAAGAACTAATPAAAAAATTATGATGGAAAGTA 471

Db 533 GTGAAGAAGAAAGTGAAGACTTTTATGAGGAGCTGATCAATPAACTATGATGAGGAAAGTC 592

Qy 472 CACGGGAGATGAGA-----ATGTGGATTATPAAATGATGAATTTTGTGAG 519

Db 593 CATGGTGAAGAAAGATGATCCCTGATCCGCTCGATATGATGATGCTGTTTTTCTGGAG 652

Qy 520 TTGGTGAATGCCCTTGGTCATATPAAATGATGATGACGATGATG-----AT 564

Db 653 TTGTCGATGCGCTGATCACTAGATGAGAGGAGGCAATGACACCTCA 712  
 QY 565 GATGAGAGATCTCGAAGAAAGAAAGCAAAAGATCTGAGATCAAGAGT 624  
 Db 713 GATGAAAGCAGATGACAGCAAGAAAGATCTGCAAGTAAAGAAAGGAGCACT 772  
 QY 625 GATTAAGAAAGCGCCACCTCG-----GAAATTTCTTCTGATTAATTTTGA 675  
 Db 773 GCTATTGAAAGCAAAAGAGTTCCAGAAAGATCTCCAAATGACATGATCTTCA 832  
 QY 676 GGCATTTCTCAATGTTTCAGATTAAGGCGCAGCAGAAAGAACTAAAGGAAATTA 735  
 Db 833 GCAATGCTCTCAATGTTCTGAGATGATGATGATGATGATGATGATGATGAT 892  
 QY 736 GAACTCAAGGAGAGAGTCCCGAGGAGATTTCTCTGATGATGATGATGATGAT 795  
 Db 893 GAGCTCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 952  
 QY 796 GAGCAAAATGCTAAATCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 855  
 Db 953 GAGCCCAATGCTCAATGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1012  
 QY 856 TGTAGGCAATGTTTAAATGATGATGATGATGATGATGATGATGATGATGAT 915  
 Db 1013 TCCCGGCGCTGCTTAAATGATGATGATGATGATGATGATGATGATGATGAT 1072  
 QY 916 TATAAGCGGAT 975  
 Db 1073 TATAAGCGGAT 1132  
 QY 976 CAGATTTGAT 1035  
 Db 1133 CTTTGTGAGAGAGAGAGAGAGAGAGAT-----GCAATGCTCAATGATCCC 1177  
 QY 1036 CCAACCAAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1095  
 Db 1178 GCGTCAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1237  
 QY 1096 AGACCCCCCAATTAATGATGATGATGATGATGATGATGATGATGATGAT 1155  
 Db 1238 AATGCTCAAGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1297  
 QY 1156 ACTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1215  
 Db 1298 AATG-----ACTGAGCTTCAAGTTCTTCAAGAGCTTAATCTGCTGATCAG 1342  
 QY 1216 AGCTCTCTGAGAGAGAGATCTGATGATGATGATGATGATGATGATGATGAT 1275  
 Db 1343 ACTCCCAAAAGAGAGAGCT---AGTCAAGCCCACTCACTGATGATGATGATGAT 1399  
 QY 1276 CCTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335  
 Db 1400 CCTCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1459  
 QY 1336 ACTTACTGAGAGATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1395  
 Db 1460 ACTTACTGAGAGATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1519  
 QY 1396 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1455  
 Db 1520 GTCTTCAAGTTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1579  
 QY 1456 GATACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515  
 Db 1580 ---AACCCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1636  
 QY 1516 ATAGAGCTGAT 1575  
 Db 1637 ATTAGCTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1696  
 QY 1576 CCAAGGAGAGCTTGTGAGAGATGATGATGATGATGATGATGATGATGATGAT 1635  
 Db 1697 CCAAG 1756

QY 1636 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1695  
 Db 1757 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1816  
 QY 1696 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1755  
 Db 1817 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1876  
 QY 1756 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1815  
 Db 1877 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1936  
 QY 1816 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1875  
 Db 1937 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1996  
 QY 1876 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1935  
 Db 1997 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2056  
 QY 1936 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1995  
 Db 2057 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2116  
 QY 1996 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2055  
 Db 2117 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2176  
 QY 2056 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2115  
 Db 2177 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2236  
 QY 2116 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2175  
 Db 2237 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2296  
 QY 2176 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2235  
 Db 2297 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2356  
 QY 2236 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2239  
 Db 2357 TTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2360

RESULT 12  
 DR762983 937 bp mRNA linear EST 21-JUL-2005  
 LOCUS HSC4.147 A11.G1 A037 NIH\_MGC\_262 Homo sapiens cDNA clone  
 DEFINITION IMAGE:7972631 5', mRNA sequence.  
 ACCESSION DR762983  
 VERSION DR762983.1 GI:71055683  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Buxarjovca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 937)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgs@bbs-remail.nih.gov  
 Tissue Procurement: BresaGen, Inc.  
 cDNA Library Preparation: Express Genomics, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 University of Georgia



Clone distribution: MGC clone distribution information can be found through the J.M.A.G.E. Consortium/JMLN at:

http://jmln.jml.gov

plate: JLM 17091 row: b column: 21

Seq primer: JENREV (CAGGAACAGCTATGACC)

High quality sequence stop: 837.

Location/Qualifiers

1. 837

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:7972631"

/sex="male"

/tissue\_type="embryonic stem"

/cell\_type="human embryonic stem cells"

/cell\_line="BG01"

/lab\_host="DH10B-T1 phage-resistant E. coli"

/clone\_id="NIH\_MGC\_262"

/note="Vector: pexprs-1; Site\_1: NotI; Site\_2: EcoRV; RNA obtained from human embryonic stem cells isolated from the inner cell mass of blastocyst stage embryos and differentiated to an early neural progenitor cell type. Cell line id and NIH Registry designation is BG01. Positive for Nestin and Musashi expression. Passage number 18. cDNA primed using oligo-dT primer: 5'-GACTAGTCTTAGATCGGAGCGGCCCT(7)25-3' and cloned into the EcoRV/NotI sites of pexprs-1. This primary library is non-normalized (normalized primary library is NIH\_MGC\_259). It was constructed by Express Genomics (Frederick, MD). Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Note: this is a Mammalian Gene Collection library."

ORIGIN

Query Match 37.0%; Score 830; DB 10; Length 837;  
Best Local Similarity 100.0%; Pred. No. 7.2e-212; Indels 0; Gaps 0;  
Matches 830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1099 ACCCCACCATTAATGTCTGGAATCAAGAATACAGACAGTATGAGGAGGAGT 1158  
8 ACCCCACCATTAATGTCTGGAATCAAGAATACAGACAGTATGAGGAGGAGT 67  
1159 GAAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAGAGAGAGAGAGT 1218  
68 GAAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAGAGAGAGAGAGT 127  
1219 TCCTCTGAAGCAAAATTCCTGCGGTCAACCAATTAAGAGAGAGAGAGAGT 1278  
128 TCCTCTGAAGCAAAATTCCTGCGGTCAACCAATTAAGAGAGAGAGAGAGT 187  
1279 CCTGAGATGTGAGTGTGCTGCTGAGAGCTCATATGTTAGAGTCTCATGGCACT 1338  
188 CCTGAGATGTGAGTGTGCTGCTGAGAGCTCATATGTTAGAGTCTCATGGCACT 247  
1339 TACTATGACAAATTTCTGTGCCATCTCTAGTTAATTTGGAGCAAAATCTGAGACAGT 1398  
248 TACTATGACAAATTTCTGTGCCATCTCTAGTTAATTTGGAGCAAAATCTGAGACAGT 307  
1399 TATGAGTTAAGTCAAGAAATCTAGCATCAAGTCTCCCGCTGAGAGAGTGTGAT 1458  
308 TATGAGTTAAGTCAAGAAATCTAGCATCAAGTCTCCCGCTGAGAGAGTGTGAT 367  
1459 ACTCTCCAGT 1518  
368 ACTCTCCAGT 427  
1519 CAGCTGAGT 1578  
428 CAGCTGAGT 487  
1579 CCGCAGCTTGTGACAGTTCTGCGCTTGTGTGATGACAGAGAGAGAGAGAGT 1638  
488 CCGCAGCTTGTGACAGTTCTGCGCTTGTGTGATGACAGAGAGAGAGAGAGT 547

Qy 1639 TGTCAATGATGATTCAGAGTGTCAAAACCGCTTTCCGGAGTCCGCTGCAAGACAGTGC 1658  
Db 548 TGTCAATGATGATTCAGAGTGTCAAAACCGCTTTCCGGAGTCCGCTGCAAGACAGTGC 607  
Qy 1699 AACACCAAGACAGTCCCGTGTCTACCTGGCTGTCCGAGAGTGTACCCCTGACCTTGTCTT 1758  
Db 608 AACACCAAGACAGTCCCGTGTCTACCTGGCTGTCCGAGAGTGTACCCCTGACCTTGTCTT 667  
Qy 1759 ACTGTGAGCCGCTGACCATTTGGGACAGTAAATGTCTCTGCAAGAACTGAGTAT 1818  
Db 668 ACTGTGAGCCGCTGACCATTTGGGACAGTAAATGTCTCTGCAAGAACTGAGTAT 727  
Qy 1819 CAGCGGGGCTTCGAGT 1878  
Db 728 CAGCGGGGCTTCGAGT 787  
Qy 1879 TTTATCAAGATCTCTGTGACAGT 1928  
Db 788 TTTATCAAGATCTCTGTGACAGT 837  
RESULT 13  
AK154565  
LOCUS  
DEFINITION  
Mus musculus MOD-derived CD1c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone: F630047M06 product: enhancer of zeste homolog 1 (Drosophila), full insert sequence.  
ACCESSION  
AK154565.1 GI:74181971  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murine; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtractions of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Komuro, H., Akiyama, J., Nishi, K., Kitanaka, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakeguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, B., Matsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, S., Arakawa, K., Irawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nishikido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabul, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barab, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bul, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,



CONSTRM	Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Nomberts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kikutsuki, S. and Hayashizaki, Y.
JOURNAL	RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium
PUBLISHED	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
REFERENCE	11217851
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, I., Otsu, K., Saito, R., Suzuki, H., Yamakawa, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bulc, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Baralov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Draganti, T.A., Fletcher, C.P., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gisel, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Malais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Q.D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yagis, C., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawaji, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
CONSTRM	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420 (6915), 563-573 (2002)
PUBLISHED	12466851
REFERENCE	
AUTHORS	
CONSTRM	Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Welle, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Bremner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, V., Allen, J.E., Ambesi-Imbondato, A., Apweiler, R., Attaraliya, R.N., Bailey, T.L., Banal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.P., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gusic, S., Hachisaka, S., Harbers, M., Hayashizaki, T., Hensch, T.K., Hirokawa, N., Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jaki, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelsio, J., Kitamura, H., Kitano, H., Kollas, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liutau, S., McWilliam, S., Madan, B., Mader, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Miyagawa, F., Miyake, S., Morita, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakanchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavoni, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Roset, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Sempke, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, R., Sugiyama, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S.,
TITLE	
JOURNAL	
PUBLISHED	
REFERENCE	
AUTHORS	
CONSTRM	Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamashita, H., Zaborovskiy, E., Zhu, S., Zimmer, A., Hilde, W., Bulc, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kato, T., Kawai, H., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawai, H., Kawasashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Pleesky, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanishi, A., Okamura, Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.
JOURNAL	PANTOM Consortium
PUBLISHED	The transcriptional landscape of the mammalian genome Science 309 (5740), 1559-1563 (2005)
REFERENCE	16141072
AUTHORS	Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakaniishi, M., Nakamura, M., Nishida, H., Yag, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hallman, J., Mattick, J.S., Hume, D.A., Lipovich, L., Pang, K.C., Hallman, J., Mattick, J.S., Hume, D.A., Lipovich, L., Baralov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlestedt, C.
CONSTRM	RIKEN Genome Exploration Research Group
TITLE	Antisense transcription in the mammalian transcriptome
JOURNAL	Science 309 (5740), 1564-1566 (2005)
PUBLISHED	16141073
REFERENCE	
AUTHORS	Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (30-Mar-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-research@riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
CONSTRM	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
COMMENT	Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/.
FEATURES	Location/Qualifiers
SOURCE	1. 4089 /organism="Mus musculus" /mol_type="mRNA" /strain="NOD" /db_xref="PANTOM DB:F630047M06" /db_xref="taxon:10090" /clone="F630047M06" /cell_type="NOD-derived CD11c +ve dendritic cells" /clone_lib="RIKEN full-length enriched mouse cDNA library" 92. 2344 /note="unlabeled protein product; enhancer of zeste homolog 1 (Proscopia) (MGI:1097695 G8 U60453, evidence: BLASTN, 99%, match=3459) putative" /codon_start=1 /protein_id="BAE32680.1" /db_xref="GI:74181972" /translation="MRKDDIASPTSKCTYKRVKSEVRLRLQKLGANMGKAL YVANPAKQKGTGLINBEWKKLVRQVQPMKRVSGHPLKKCALISIPGDSQDMML



QY 2056 TTGGCAATCATTCGGTAATCCAACTGCTATGCAAAAGTATGATGTTAAGCGTAT 2115  
 DB 2159 TTGGCAACCATTCAGTGAACCCCACTTTATGCAAGTGTATGTTAATGAGAT 2218  
 QY 2116 CACAGATAGTATTTTCCCAAGAGAGCCATCCAGACTGGCGAAGAGCTGTTTTTGTAT 2175  
 DB 2219 CACCGCATGGATCTTTCCCAAGAGAGCAATTCAGGCTGGCGAAGAGCTCTTTTGTAT 2278  
 QY 2176 TACAGATACAGCAGCGCTATGCTCCCTGAGTATGTCGGATGGAAGAAATGGAATC 2235  
 DB 2279 TTAGGTACAGCCAGCTGATGCTCCCTCAAGTATGTGGCATCGAAGAGGAAACGACGCTC 2338

RESULT 14  
 LOCUS AK140694 4194 bp mRNA linear HTC 21-SEP-2005  
 DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
 enriched library, clone:B930092M06 product:enhancer of zeste  
 homolog 1 (Drosophila), full insert sequence.  
 ACCESSION AK140694.1 GI:74180279  
 VERSION AK140694.1  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS 2  
 TITLE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS 3  
 TITLE Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Oyake, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multiplexed sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS 4  
 TITLE Kawai, T., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,  
 Saito, K., Okazaki, Y., Gojobori, T., Bono, H., Kasekawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
 Kuell, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,  
 Quackenbush, J., Schriml, L.M., Staudt, J., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Kono, H.,  
 Baldarelli, R., Barsh, G., Blake, J., Botfield, D., Bojunga, N.,  
 Carinini, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Hill, D.,  
 Fleischer, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
 Mochizuki, L., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
 Toyo-Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,  
 Wyszynski, B., Yoshida, K., Hasegawa, Y., Kawai, H., Koltz, S.,  
 Hayashizaki, Y.,  
 RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium

FUNCTIONAL ANNOTATION OF A FULL-LENGTH MOUSE CDNA COLLECTION  
 Nature 409 (6821), 685-690 (2001)  
 11217851

REFERENCE  
 AUTHORS 5  
 TITLE Okazaki, Y., Furuno, M., Kasekawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Oshio, N., Saito, R., Suzuki, H., Yamana, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K.W., Blake, J.A., Bridt, D., Brusic, V.,  
 Chotlov, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S.,  
 Guelinich, S., Hirokawa, N., Jackson, J., Jarvis, E.D., Kanai, A.,  
 Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Kongaya, A.,  
 Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Maglott, D.R.,  
 Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, M.J., Petrea, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Sempole, C.A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L.G., Wyszynski, B., Yang, Y., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imokani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 PANTOM Consortium  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420 (6915), 563-573 (2002)  
 12466851

CONSRMT  
 TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 6  
 Carninci, P., Kasekawa, T., Katayama, S., Gough, J., Fritch, M.C.,  
 Maeda, N., Oyama, R., Ravasi, T., Lennard, B., Wells, C., Kodat, R.,  
 Shimokawa, K., Bai, J., Brenner, S.E., Batalov, S., Forrest, A.R.,  
 Zavolan, M., Davis, M.J., Wilming, L.G., Adkins, V., Allen, J.E.,  
 Amesell, Impombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L.,  
 Banal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M.,  
 Chiu, K.P., Choudhary, V., Christofels, A., Clutterbuck, D.R.,  
 Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G.,  
 di Bernardo, D., Down, T., Engstrom, P., Fegolotti, M., Faulkner, M.,  
 Fletcher, C.F., Fukuhima, T., Furuno, M., Futaki, S., Gariboldi, M.,  
 Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.B.,  
 Guelinich, S., Harbers, M., Hayashi, Y., Henrich, T.K., Hirokawa, N.,  
 Hill, D., Hummel, L., Iacono, M., Ieko, K., Iwama, A., Ishikawa, T.,  
 Jaki, M., Kanapin, A., Katoh, M., Kawasawa, Y., Keiso, J., Kitamura, H.,  
 Kitanaka, H., Kollas, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K.,  
 Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J.,  
 Lin, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L.,  
 Matsuda, H., Matsuzawa, S., Miki, H., Mignone, P., Miyake, S.,  
 Morris, K., Moutreuil, Taber, S., Mulder, N., Nakano, N., Nakachi, H.,  
 Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O.,  
 Okazaki, Y., Orlowski, V., Pang, K.C., Pavan, W.J., Pavoni, G.,  
 Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z.,  
 Ringwald, M., Ros, B., Ruan, Y., Salzberg, S.L., Sandelin, A.,  
 Schneider, C., Schonbach, C., Sekizuchi, K., Sempole, C.A., Sene, S.,  
 Sees, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D.,  
 Sincalir, B., Sperling, S., Stupka, S., Sugita, K., Sultana, R.,  
 Takenaka, Y., Taki, K., Tamojia, K., Tan, S.L., Tang, S., Taylor, M.S.,  
 Tegen, J., Teichmann, S.A., Ueda, H., van Nimwegen, E., Verardo, R.,  
 Wei, C.L., Ye, K., Yamashita, H., Zdobych, B., Zhu, S., Zimmer, A.,  
 Hide, W., Bult, C., Grimond, S.M., Teasdale, R.D., Liu, B.T.,  
 Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A.,  
 Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, Katayama, M.,  
 Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M.,  
 Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M.,  
 Kondo, S., Kono, H., Nakano, K., Niimura, N., Nishio, T., Okada, M.,  
 Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K.,

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS  
7  
Katsuyama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Fritch, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Motegani-Tabar, S., Liang, Z., Lenhard, B. and Wahlstedt, C.  
CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS  
8  
Riken Genome Exploration Research Group  
Science 309 (5740), 1564-1566 (2005)  
16141073  
8 (bases 1 to 4194)  
Arkawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y.  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/.  
URL: http://fantom.gsc.riken.jp/.  
Location/Qualifiers  
1. 4194  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM:DB:B930092M06"  
/db\_xref="taxon:10090"  
/clone="B930092M06"  
/tissue\_type="cerebellum"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
119..2371  
/note="unmated protein product; enhancer of zeste homolog 1 (Protophila) (MCD|MG1:1097695|GB|NM\_007970, evidence: BLASTN, 98%, match=4156)  
putative"  
/codon\_start=1  
/protein\_id="BAE24446.1"  
/db\_xref="GI:74180280"  
/translation="MRKNDIASPTSKITTYWKRKVSERYRLAQKRLQANMGAKL  
YVNPFAVQERTQILNEWRKLAVQVPMKPVSGHPPLKCTTISTIFPGDSQMLM  
RSLNTVALVPIWTSWPLQONFVDETVLNPIMGDEVBEDETIFELINNYDK  
VHGEEMIKGSVLISDAVLELVALNYSDBEKGNDPDKQDQSDNALPTPKR  
KRHA1EGNKSSKQFPMDMFSALASFPENGVDDMKERYRELTEKSDNALPPOC  
TPNTDGPVAKSVOROSLHSPHTLFCRCFPCYDCITLPHFHTLPVYKXKEITLPE  
PCGTDCPLILGAKRYATLNPBRSKCRRRRRHRYVSASSNSASAMATKCEIDP  
RDTGNDWASSSSSEANSRCQPTPKQASAPAPQOLCVESAPSPVENTGSESLFVTHG  
TYPNFNCIARLITKCKQVFOFAYVESLILKLPTEILAMPQKKRKRLLMAHCR  
KIQLKDNNSYOVNVYQPCDHPBRCDSCTCMTQNFCEKFCSCSPCONRPFGRCR

Query Match 37 0%; Score 828.8; DB 6; Length 4194;  
Best Local Similarity 63.6%; Pred. No. 2.5e-211;  
Matches 1413; Conservative 0; Mismatches 732; Indels 75; Gaps 7;

55 GTAAATCAAGTACATGCACTGAGACAGCTCAAGGTTTCAAGCAAGCTGATGAAGTA 114  
182 GTGAATCTGATGATATATGCGGCTTTCGACACTCAAAAGGCTCCAGGCAAAATATGGAGCA 241  
115 AAGATATGTTTATGTTCCATTCGTCAAGAAATTTGGAAAGAACGAAATCTTAAACCA 174  
242 AAGGCTGTATGTGGCAAAATTTTGCAGAAAGTTCAAGAAATCCAAATCTTATGAA 301  
175 GAATGGAAGACGCAAGATATACAGCTGTGCAATCTTGAATTTGAGCTCATTTGCGC 234  
302 GAGTGAAGAAATCTGTGTCAGCTGTTCAGCAATGAGCCGTGAGTGGCAACCT 361  
235 GGAATGAGGAGTTC--GGTACCAATGACTTGATTTTCCAAACAATCATCCCA 291  
362 TTTCTGAAAAAGTACATATAGAGATATTTCCAGGGTTCGACACGAGATATTTG 421  
292 TTAAAGCTGATGATGAGTTGCTTCAATATATATATATATATATATATATATATATAT 351  
422 ATGCGGTCTGAACTGTTGCACTGTTTCCATCATGATATCTGATCTTCCCTTCAG 481  
352 CAGATTTTATGTGGAAGATGAACTGTTTATCATTAATCTTATATATGAGATGAA 411  
482 CAGATTTTATGTGGAAGATGAGCGTTTGTGCAATATTCATATCATGAGTGAAGAG 541  
412 GTTTTATGATGATGATGATCTTCTTATGAAAGAACTATTAATAATATATATGAGAAAGTA 471  
542 GTGAAGGAAGAGATGAGATCTTCTATGAAAGCTGATCATATATATATATATATATATAT 601  
472 CAGCGGATGAGAG-----ATGCGGTTTATATGATGAAATTTTGTGAG 519  
602 CAGGTGAAGAGAGATGATCTTCTGATCTGTGATGAGATCTGTGTTTCTGAG 661  
520 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 579  
662 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
580 GAGAAAG 615  
722 GATGAAAG 781  
616 CAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 675  
782 GCTATGAG 841  
676 GCCATTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735  
842 GCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 901  
736 GAATCAACGAG 795  
902 GAGCTGAGAGAGATGAT 961  
796 GAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 855  
962 GAGCCCAAG 1021  
856 TGTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915  
1022 TGCCGGGCGTGTGTTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081  
916 TATAGGAG 975  
1082 TATAGGAG 1141  
976 CAGATTTGAG 1035  
1142 CTTTGTGAG 1186  
1036 CCAACAAACGTCAG 1095  
1187 CGGTCAAG 1246

```

Oy 1096 AGACCCCAACATTAATGTCTGGAATCAAGATACAGACATGATAGGAGACAGG 1155
Db 1247 AATGATCAGCTTCTGCTATGCTGTAATAAGAGAGACATGATAGAACATCTGC 1306
Oy 1156 ACTGAAACGGGGGAGAGAAATGATTAAGAAAGAAAGAAAGAAATGAAATCTTG 1215
Db 1307 AATG-----ACTGGGCTCAGTCTTCTGAGAGGCTAATCTCGCTGACG 1351
Oy 1216 AGCTCTCTGAGCAAAATCTCGGTCTCAAAACCAATTAAGTGAAGCAATATGAA 1275
Db 1352 ACCCCACGAAACAGAAAGC---CAGTCCAGCCCAAGCTCAGCTGTGTGTGAAAGC 1408
Oy 1276 CCTCTGAGAAATGTGAGTGAAGTGTGCTGAAGCTCAATGTTAGAGTCTCATGTGC 1335
Db 1409 CCTCAGAGCCGCTGAGATGACCGAGCCGAAAGATCTCTTTCCGAGTCTTCCAGGC 1468
Oy 1336 ACTTATGATGACAAATTTCTGTGCAATGCTAGATTATTTGAGACCAAAATGTAACG 1395
Db 1469 ACCATTTTCAACAATCTGTCTCAATAGCCAGCTTCTGGGACAAAGACATGCAAGCAG 1528
Oy 1336 GTGTATGAGTTTGAAGTCAAAATCTAGATCATAGCTCCAGCTCCGCTGAGATG 1455
Db 1529 GTCTTCAAGTTTGCAGTCAAGATCACTTATCTTAAGCTCCAAAGATAGCTCAG 1588
Oy 1456 GATACCTCTCAAGGAAAGAAAGAGAAACACCGGTTGTGGGCTGCACTGCAAGAA 1515
Db 1589 ---AACCTTGCACAGAGAGAAAGAAAGAAACAGGTGTGTGGCCGACACTGCAAGAA 1645
Oy 1516 ATACAGCTGAAAGAGACCGCTCTCTTAACAATGTTTACAATCAACCTGTGATCAT 1575
Db 1646 ATTCAGCTGAAAGAAAGATTAACAATCTTAACAAGTATTAACAACCACTGTGACAC 1705
Oy 1576 CCAAGGACGCTTGTGACAGTTCGTGCTGTGTGTGATGACCAAAATTTTGTGAAG 1635
Db 1706 CCAAGCCGCTGTGTGACAGCAATGCCCCCTCATATGACCCAGAACTTTGTGAAG 1765
Oy 1636 TTTTGTCAATGTATGTGACAGTGTCAAAACCGCTTTCGGGAGTCCGCTGCAAGCAG 1695
Db 1766 TTCTCCAGTGAAGCCCAAGCTGCAAGATGCTTCTGTGTGTGTGATGACTCAG 1825
Oy 1636 TGCACACCCAGACAGTCCGCTGTCTAAGCTGTCCGAGATGTGACCTGACCTGT 1755
Db 1826 TGCATATCAACAGCAATGTCATAGTCTAAGTGTGAGTGTGACCTGACCTGTGAC 1885
Oy 1756 CTACTTGTGAGCCGCTGACCAATTTGGAGCAATTAATGTCTCTGCAAGACCTGAC 1815
Db 1886 CTCACCTGTGGGCTCAGAGCACTGGACATGTAAAGTGTGTCTTGTGCAAAACCTGAC 1945
Oy 1816 ATTGACGGGGGCTCCAAAGAGCATGATTTGTGGAACAATGAGTGGCAGGCTGGGG 1875
Db 1946 ATCCAGGCTGGCTCAAAAGAGCACTGTGCTGGCCCTTCCGATGTGGCCGAGTGGGC 2005
Oy 1876 ATTTTATCAAAAGATCTGTGCAAAATGAAATGAAATCATCTCAGAAATCTGTGAGAG 1935
Db 2006 ACCCTCATCAAGAGATCTGTGCAAAAGATGAAATCATCTTGTGAAATTTGTGTAGCTC 2065
Oy 1936 ATTTTCTCAAGATGAGCTGACAAAGAGGAAAGTGTATGATTAATACATGTGACGCTT 1995
Db 2066 ATCTCTCAGATGAGAGGCTGATGTGAGGGAAGGCTATGATTAATACATGTGACGCTT 2125
Oy 1996 CTGTTCAACTGAAAGATTTTGTGTGATGACCCGCAAGGTTAAACAAATTTG 2055
Db 2126 CTCTTCAACTCAACAGATTTTGTGTGATGACCCGCAAGGTTAAACAAATTTG 2185
Oy 2056 TTTGCAAAATCATTCGTTAATCAAACTGCTATGCAAAAGTATGATGTTAACGGTAT 2115
Db 2186 TTTGCAAAACATTCAGTGAAGCCCAAGCTTATGCAAAAGTGTATGTTGATGAGAGAT 2245
Oy 2116 CACAGATGAGTATTTTGTGCAAGAGGCAATCCAGCTGCGGAAGAGCTGTTTTTGTAT 2175
Db 2246 CACCCCATTTGGGATCTTTGTGCAAGAGAGCAATTCAGGCTGGGAAGAGCTCTCTTTGAT 2305
Oy 2176 TACAGATACAGCCAGGCTGATGCTGGAAGTATGTCTGGCATGGAAGAAATGGAATC 2235

```

```

Db 2306 TATAGTACAGCCAGCTGATGCCCCCTCAAGTATGTGGCATGCAAGGGAACGAGCTC 2365

RESULT 15
LOCUS CX783825
DEFINITION HSC3 26 B01.g1 A036 NIH_MGC 260 Homo sapiens cDNA clone
IMAGE:7478092.5', mRNA sequence.
ACCESSION CX783825
VERSION CX783825.1 GI:58300615
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 840)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgs@bse-remail.nih.gov
Tissue Procurement: BresaGen, Inc.
cDNA Library Preparation: Express Genomics, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU)
DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
University of Georgia
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNU at:
http://image.llnl.gov
Plate: LILM15792 row: 1 column: 02
Seq primer: JENREV (CAGGAACAGCTATGACC)
High quality sequence stop: 840.
Location/Qualifiers
1. 840
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7478092"
/sex="male"
/tissue_type="human embryonic stem cells"
/cell_type="human embryonic stem cells"
/cell_line="B01"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_id="NIH_MGC_260"
/name="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV.
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos. Cell line
id and NIH Registry designation is B01. Positive for
SSR33, SSR44, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and
Nanog expression; negative for SSR1 expression. Passage
number 21. cDNA primed using oligo-dT primer:
5'-pGACTGTTCTAGATCGCAGCGGCGCCCTT)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. This primary library
is non-normalized (normalized primary library is
NIH_MGC_261). It was constructed by Express Genomics
(Fredrick, MD). Sequence ends have been trimmed to
exclude vector and regions below Phred quality 16. Note:
this is a Mammalian Gene Collection library."

```

## ORIGIN

```

Query Match 36.9%; Score 826.4; DB 9; Length 840;
Best Local Similarity 99.6%; Pred. No. 6.7e-211;
Matches 838; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
Oy 1040 CAAAAGTCCAGAGGCGCGCAGAAAGAGACGGCTTCCCAATTAACATGACAGGCGCCAGCA 1099
Db 1 CCAAAGTCCAGAGGCGCGCAGAAAGAGACGGCTTCCCAATTAACATGACAGGCGCCAGCA 60

```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2006, 11:22:29 ; Search time 442 Seconds  
(without alignments)  
9486.776 Million cell updates/sec

Title: US-10-773-302-1\_COPY\_90\_2330

Perfect score: 2241

Sequence: 1 atggggccagactgggaagaa.....gagaatggaatcccttga 2241

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1\_COMB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5\_COMB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6a\_COMB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6b\_COMB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7\_COMB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H\_COMB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP\_COMB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP\_COMB.seq:\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP\_COMB.seq:\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfile1.seq:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2241	100.0	2600	3	US-09-589-892B-1
2	1993.8	89.0	2476	3	US-10-104-047-1192
3	218.8	9.8	489	3	US-09-589-892B-5
4	148.4	6.6	1406	3	US-09-699-266A-10
5	138.2	6.2	3148	3	US-09-699-266A-8
6	135.4	6.0	2136	3	US-09-177-249-1
7	135.4	6.0	2136	3	US-09-812-283-1
8	135.4	6.0	2136	3	US-09-071-838A-1
9	135.4	6.0	2136	3	US-09-061-769A-1
10	135.4	6.0	2799	3	US-10-176-884-16
11	119.2	5.3	260	3	US-09-640-211A-349
12	119.2	5.3	260	3	US-09-640-211A-591
13	118.4	5.3	2322	3	US-09-061-769A-4
14	115	5.1	2327	3	US-09-061-769A-3
15	70.4	3.1	659	3	US-09-270-767-765
16	70.4	3.1	659	3	US-09-270-767-16047
17	68.8	3.1	2652	3	US-09-248-796A-4828
18	63	2.8	7218	2	US-08-232-463-14
19	62.2	2.8	12078	2	US-08-799-451-835
20	62.2	2.8	12308	3	US-09-854-133-422
21	55	2.5	2409	3	US-09-614-221A-204
22	53.6	2.4	767677	3	US-09-949-016-12147
23	53.6	2.4	767677	3	US-09-949-016-17361

24	53.4	2.4	31391	3	US-09-949-016-14319	Sequence 14319, A
25	52.6	2.3	11907	2	US-08-061-376-4	Sequence 4, Appli
26	52.6	2.3	14255	2	US-08-320-559-1	Sequence 1, Appli
27	52.6	2.3	14255	2	US-08-327-392-1	Sequence 1, Appli
28	52.6	2.3	14255	2	US-08-306-618-55	Sequence 35, Appli
29	52.6	2.3	14255	2	US-08-545-860D-1	Sequence 1, Appli
30	52.6	2.3	14255	2	PCT-US94-04496-1	Sequence 9123, Ap
31	52.4	2.3	3102	5	US-10-131-831-9123	Sequence 19856, S
32	51.6	2.3	601	3	US-09-949-016-19856	Sequence 1, Appli
33	51	2.3	640681	3	US-09-790-988-1	Sequence 4970, Ap
34	50.4	2.2	1671	3	US-09-248-796A-4970	Sequence 270, App
35	50.4	2.2	3217	4	US-10-094-749-270	Sequence 6, Appli
36	50.4	2.2	7015	4	US-09-177-249-6	Sequence 6, Appli
37	50.4	2.2	7015	5	US-09-812-283-6	Sequence 6, Appli
38	50.4	2.2	7015	5	US-09-071-838A-6	Sequence 6914, Ap
39	49.4	2.2	576	3	US-09-489-039A-6914	Sequence 4666, Ap
40	48.8	2.2	876	3	US-09-248-796A-4666	Sequence 1, Appli
41	48.8	2.2	1452	3	US-09-876-221-1	Sequence 2, Appli
42	48.6	2.2	15225	2	US-08-892-403A-2	Sequence 10, Appli
43	48.6	2.2	15225	2	US-09-291-894-2	Sequence 1, Appli
44	48.6	2.2	15225	3	US-09-827-688-10	Sequence 1, Appli
45	48.4	2.2	10304	3	US-09-627-465B-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-589-892B-1  
Sequence 1, Application US/09589892B  
Patent No. 6689583  
GENERAL INFORMATION:  
APPLICANT: Jenuwein, Thomas  
APPLICANT: Laible, Gotz  
APPLICANT: O'Carroll, Donal  
APPLICANT: Eisenhaber, Frank  
APPLICANT: Rea, Stephen  
TITLE OF INVENTION: Chromatin-Regulator Genes  
FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/09/589, 892B  
CURRENT FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945,988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
PRIOR FILING DATE: 1995-05-10  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(89)  
FEATURES:  
NAME/KEY: CDS  
LOCATION: (90)..(2330)  
FEATURES:  
NAME/KEY: 3'UTR  
LOCATION: (2331)..(2600)  
US-09-589-892B-1

Query Match 100.0%; Score 2241; DB 3; Length 2600;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCTGGAGAAATCTGAGAGGACCACTTTGTGGCCGGAAGCGTTAA 60  
DB 90 ATGGCCGAGCTGGAGAAATCTGAGAGGACCACTTTGTGGCCGGAAGCGTTAA 149  
QY 61 TCAAGTATCATGCGAGCTGAGACAGCTCAAGAGGTTCAAGAGCGTGAATAAGAGT 120





RESULT 2  
US-10-104-047-1192  
; Sequence 1192, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patemlin Ver. 2.1  
; SEQ ID NO 1192  
; LENGTH: 2476  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1192

Query Match 89.0%; Score 1993.8; DB 3; Length 2476;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 2; Indels 117; Gaps 1;

QY 1 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAAGCGTGTAAA 60  
DB 93 ATGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAAGCGTGTAAA 152  
QY 61 TCAGAGTCAATGCGAAGTCAAGAGCTTCAAGAGCTGATGATGAAGAT 120  
DB 153 TCAGAGTCAATGCGAAGTCAAGAGCTTCAAGAGCTGATGATGAAGAT 212  
QY 121 ATGTTAAGTCCATCGTCAGAAATTTTGGAAAGACGSAATCTTAAACCAAGATGG 180  
DB 213 ATGTTAAGTCCATCGTCAGAAATTTTGGAAAGACGSAATCTTAAACCAAGATGG 272  
QY 181 AAACGCGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTTGGCGGAGCT 240  
DB 273 AAACGCGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTTGGCGGAGCT 332  
QY 241 AGGAGTGTGGGTACCAAGTCACTTGATTTTCCAACAAGTATCCCATTAAGACT 300  
DB 333 AGGAGTGTGGGTACCAAGTCACTTGATTTTCCAACAAGTATCCCATTAAGACT 337  
QY 301 CTGATGAGTTCCTTCAAGTACCAATATGATTTCTGGTCTCCCTACAGAGATTTT 360  
DB 338 ----- 337  
QY 361 ATGTTGAGAGATGAATCTGTTTACATTAATCTTATATGAGAGATGAAGTTTAAAT 420  
DB 338 --GGTGAAGATGAATCTGTTTACATTAATCTTATATGAGAGATGAAGTTTAAAT 395  
QY 421 CAGAGTGAATCTTCAATGAAGAACTAATTAATAATTAATGATGAGAAAGTACACGGGAT 480  
DB 396 CAGAGTGAATCTTCAATGAAGAACTAATTAATAATTAATGATGAGAAAGTACACGGGAT 455  
QY 481 AGAGATGTGGTATTAATAATGATGAATTTTGTGAGATTGGTGAATGCCCTTGGTCA 540  
DB 456 AGAGATGTGGTATTAATAATGATGAATTTTGTGAGATTGGTGAATGCCCTTGGTCA 515  
QY 541 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 516 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575  
QY 601 AAAGATCTGAGAGATCAACGAGATGATGAATGAAGAAACCGCCCACTCCGAAATTTCT 660  
DB 576 AAAGATCTGAGAGATCAACGAGATGATGAATGAAGAAACCGCCCACTCCGAAATTTCT 635  
QY 661 GATAAAATTTTGAAGCATTTCTCAATGTTTCCAGATTAAGGAGACAGCAGAGAACTA 720  
DB 636 GATAAAATTTTGAAGCATTTCTCAATGTTTCCAGATTAAGGAGACAGCAGAGAACTA 695

QY 721 AAGAAAAATATTAAGAACTCAGCAGACGAGCTCCAGGCGCACTTCTCTGTAATGT 780  
DB 696 AAGAAAAATATTAAGAACTCAGCAGACGAGCTCCAGGCGCACTTCTCTGTAATGT 755  
QY 781 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGACAAAGCTTACATCC 840  
DB 756 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGACAAAGCTTACATCC 815  
QY 841 TTTCAATGCTTTTGTGAGGCGATGTTTAAATATGATGCTGTTCTTACATCTTTTCA 900  
DB 816 TTTCAATGCTTTTGTGAGGCGATGTTTAAATATGATGCTGTTCTTACATCTTTTCA 875  
QY 901 GCAACACCAACACTTATTAAGCGGAAGAACAGAAACAGCTCTGACAAACAACTGT 960  
DB 876 GCAACACCAACACTTATTAAGCGGAAGAACAGAAACAGCTCTGACAAACAACTGT 935  
QY 961 GCAACACAGTGTATCAGACATTTGAGGAGCAAAAGATTTGCTGCTCTCAACGCT 1020  
DB 936 GCAACACAGTGTATCAGACATTTGAGGAGCAAAAGATTTGCTGCTCTCAACGCT 995  
QY 1021 GAGCGATTAAGACCCCAACAAACGTCAGAGGCCCGAGAGAGACGCTTCCCAAT 1080  
DB 996 GAGCGATTAAGACCCCAACAAACGTCAGAGGCCCGAGAGAGACGCTTCCCAAT 1055  
QY 1081 AACGATGACAGGCGCCAGACCCCAACATTAATGTCGGAATCAAAAGATACAGACGT 1140  
DB 1056 AACGATGACAGGCGCCAGACCCCAACATTAATGTCGGAATCAAAAGATACAGACGT 1115  
QY 1141 GATAGGAG 1200  
DB 1116 GATAGGAG 1175  
QY 1201 AAGATGAACCTTCAGGCTCTCTGAGCAAAATTCGAGTCAAAACCAATTAAGATG 1260  
DB 1176 AAGATGAACCTTCAGGCTCTCTGAGCAAAATTCGAGTCAAAACCAATTAAGATG 1235  
QY 1261 AAGCAATATTTGAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1236 AAGCAATATTTGAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295  
QY 1321 AAGTCTCATTTGAGCACTTACATTAATGATTTCTGTCATTTGCTAGGTTAATGGACC 1380  
DB 1296 AAGTCTCATTTGAGCACTTACATTAATGATTTCTGTCATTTGCTAGGTTAATGGACC 1355  
QY 1381 AAAACATGTAGACAGGTGTATGATTAAGTCAAAAGATCTAGCATATAGCTCAGCT 1440  
DB 1356 AAAACATGTAGACAGGTGTATGATTAAGTCAAAAGATCTAGCATATAGCTCAGCT 1415  
QY 1441 CCGGCTGAGAGATGAGATATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB 1416 CCGGCTGAGAGATGAGATATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1475  
QY 1501 GCAACCTGAGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
DB 1476 GCAACCTGAGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535  
QY 1561 CAACCTGTGATCAATCCAGCAGCTTGTGACAGTTGTCGCTTGTGTGATGACACA 1620  
DB 1536 CAACCTGTGATCAATCCAGCAGCTTGTGACAGTTGTCGCTTGTGTGATGACACA 1595  
QY 1621 AATTTTGTGAG 1680  
DB 1596 AATTTTGTGAG 1655  
QY 1681 CGCTGCAAG 1740  
DB 1656 CGCTGCAAG 1715  
QY 1741 GACCTGACCTCTGTCTTATCTTGTGAGAGCGCTGACCAATTTGGAGACAGTAAATGTGTC 1800  
DB 1716 GACCTGACCTCTGTCTTATCTTGTGAGAGCGCTGACCAATTTGGAGACAGTAAATGTGTC 1775

QY 1801 TGCAAGAACTGACGATTTACGCGGGCTCCAAAAAGCATCTATTGCTGCGACCACTTGAC 1860  
| | | | |  
DB 1776 TGCAAGAACTGACGATTTACGCGGGCTCCAAAAAGCATCTATTGCTGCGACCACTTGAC 1835  
| | | | |  
QY 1861 GTGCAAGCTGGGGGATTTTATCAAAAGATCTGTGCGAAGAAATTTGATCTCGAA 1920  
| | | | |  
DB 1836 GTGCAAGCTGGGGGATTTTATCAAAAGATCTGTGCGAAGAAATTTGATCTCGAA 1895  
| | | | |  
QY 1921 TACTGTGAGAGATTTATCTCAAGATGACCTGACAGAAAGGAAAGTATGATGATA 1980  
| | | | |  
DB 1896 TACTGTGAGAGATTTATCTCAAGATGACCTGACAGAAAGGAAAGTATGATGATA 1955  
| | | | |  
QY 1981 TACATGTGACCTTTCTGTCACTTGAAACAATGATTTTGTGTGATGCAACCGCGAAG 2040  
| | | | |  
DB 1956 TACATGTGACCTTTCTGTCACTTGAAACAATGATTTTGTGTGATGCAACCGCGAAG 2015  
| | | | |  
QY 2041 GGTAAACAAATTCGTTTGGCAATTCATCGGTAAATCCAACTGTATGCAAAAGTTATG 2100  
| | | | |  
DB 2016 GGTAAACAAATTCGTTTGGCAATTCATCGGTAAATCCAACTGTATGCAAAAGTTATG 2075  
| | | | |  
QY 2101 ATGTTAAGCGTATCAAGATAGTATTTTGGCAAGAGCCATCCAGACTGGCGAA 2160  
| | | | |  
DB 2076 ATGTTAAGCGTATCAAGATAGTATTTTGGCAAGAGCCATCCAGACTGGCGAA 2135  
| | | | |  
QY 2161 GAGCTGTTTTTGAATTACAGATACAGCGGCTGATGCTTGAAAGTATGTGGCATCGAA 2220  
| | | | |  
DB 2136 GAGCTGTTTTTGAATTACAGATACAGCGGCTGATGCTTGAAAGTATGTGGCATCGAA 2195  
| | | | |  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
| | | | |  
DB 2196 AGAGAAATGAAATCCCTTGA 2216  
| | | | |

## RESULT 3

US-09-589-892B-5

Sequence 5, Application US/09589892B

Patent No. 6689583

GENERAL INFORMATION:

APPLICANT: Jenwein, Thomas

APPLICANT: Laible, Gotz

APPLICANT: O'Carroll, Donal

APPLICANT: Riesenaber, Frank

APPLICANT: Rea, Stephen

TITLE OF INVENTION: Chromatin-Regulator Genes

FILE REFERENCE: 0652.1670001

CURRENT APPLICATION NUMBER: US/09/589,892B

CURRENT FILING DATE: 2000-06-09

PRIOR APPLICATION NUMBER: US 08/945,988

PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: PCT/EP96/01818

PRIOR FILING DATE: 1996-05-02

PRIOR APPLICATION NUMBER: DE 195 16 776.7

PRIOR FILING DATE: 1995-05-10

NUMBER OF SEQ ID NOS: 21

SOFTWARE: Patent version 3.2

SEQ ID NO 5

LENGTH: 489

TYPE: DNA

ORGANISM: Homo sapiens

US-09-589-892B-5

Query Match

Best Local Similarity 9.8%; Score 218.8; DB 3; Length 489;

Matches 265; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1756 CTATCTTGTGAGCGCGCTGACCATTTGGGACAGTAAATGTCTCTGCAAGAACTGCGAT 1815  
| | | | |  
DB 2 CTACACCTGTGGGGCTCTAGAGCACTGGGAGTGGTTCCTGTAAATACTGCGAC 61  
| | | | |  
QY 1816 ATTGACGGGGGCTCCAAAAGCATCTATTGCTGCGACCATCTGACGTGGGAGGCTGGGG 1875  
| | | | |  
DB 62 ATCCAGCGGTGACTTAAGAGCACTGCTGCGGCCCTCTGATGTGGCGGATGGGGC 121  
| | | | |

QY 1876 ATTTTATCAAGATCTGTGCGAAGAAATGAAATTCATCTCAGAAATACGTGAGAGATT 1935  
| | | | |  
DB 122 ACCCTCAATTAAGAGGTCTGCGAAGAAAGAAATTCATTTCTGAATACGTGTGAGACTC 181  
| | | | |  
QY 1936 ATTTCTCAAGATGAGCTGACAGAAAGGAAAGTATGATTAATCATGTGCGACTTT 1995  
| | | | |  
DB 182 ATCTCTAAGATGAGCTGATGACGCGGAAAGGTCTATGACAAATACATGTCCAGCTTC 241  
| | | | |  
QY 1996 CTGTTCACTTGAACAAATGATTTTGTGTGATGCAACCGGAGGTTAAACAAATTCGT 2055  
| | | | |  
DB 242 CTCTTCAACCTCAATATGATTTTGTGTGATGCTACTCGAAGAAAGGAAACAAATTCGA 301  
| | | | |  
QY 2056 TTGCAAAATCTTCGTTAAATCCAACTGCTATGCAAAAGTT 2097  
| | | | |  
DB 302 TTGCAAAATCTTCGTTAAATCCAACTGCTATGCAAAAGTT 343  
| | | | |

## RESULT 4

US-09-266A-10

Sequence 10, Application US/09699266A

Patent No. 6559354

GENERAL INFORMATION:

APPLICANT: Ma, Hongchang

APPLICANT: Morakinyo, Layo O.

APPLICANT: Odeh, Joan T.

APPLICANT: Orozco Jr., Emil M.

APPLICANT: Rafaleki, J. Antoni

TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS

FILE REFERENCE: B1164 US NA

CURRENT APPLICATION NUMBER: US/09/699,266A

CURRENT FILING DATE: 2000-10-27

PRIOR APPLICATION NUMBER: PCT/US99/08385

PRIOR FILING DATE: 1999-04-16

PRIOR APPLICATION NUMBER: 60/083,212

PRIOR FILING DATE: 1998-04-27

NUMBER OF SEQ ID NOS: 13

SOFTWARE: Microsoft Office 97

SEQ ID NO 10

LENGTH: 1406

TYPE: DNA

ORGANISM: Trilicium aestivum

FEATURES:

NAME/KEY: unsure

LOCATION: (513)

OTHER INFORMATION: n=a, c, g, or t

US-09-699-266A-10

Query Match

Best Local Similarity 5.6%; Score 148.4; DB 3; Length 1406;

Matches 390; Conservative 0; Mismatches 287; Indels 30; Gaps 4;

QY 1524 GAAAAGGACCGCTCTTAACATGTTTACAACTATCAACCTGTGATCATCCAGCA 1583  
| | | | |  
DB 303 GAAAAGGATGATGGAAGCAATGTGACAGACATGATCACATGTGGGTCCAGCA 362  
| | | | |  
QY 1584 GCCTTGTGACAGTTCGTCCTTGTGTGATGACCAAAATTTTGTGAAAAGTTTGTGCA 1643  
| | | | |  
DB 363 AATGTGCAACAAAATTTGCCCTGTGTGAAAATGAGCAATGTGAGAAATCTGTGG 422  
| | | | |  
QY 1644 ATGTAGTTCAAGATGTCAAAACCGCTTCCGGATGCC--GCTGCAAGACAGATGCAAC 1701  
| | | | |  
DB 423 GTGTTCAAAAGCTGCAAAACAGATTTAGAGCTGTGATGTCATTTGCAAAATCATGTGCA 482  
| | | | |  
QY 1702 ACCAAGACATGCCGTGTACTGTGCTGTGCGAGATGTGACCTGTGACCTGTGTC----- 1756  
| | | | |  
DB 483 AGCAGGGAATGCCATGTTTGTGCTGCATGCGGAATGTGATCCGGAATGTTTGCAGAAA 542  
| | | | |  
QY 1757 -----TTACTTGTGAGCCGCTGACCATTTGGG-----CAGTAAAAATGTGTCTG 1802  
| | | | |  
DB 543 CTGCTGGGGTGTGAGTGTGAGTGTTCACCTAGGCGAGCCACAAAGAAAGGTGATGT 602  
| | | | |  
QY 1803 CAAGAACTGAGTATTAAGCGGGGCTCCAAAAGCATC-----TATTGCTGCGACC 1853  
| | | | |

Db 603 TACCCATGCGGAAACATGAGCTCTCTTAAACAAACAAAGATTTGCTTGAAA 662  
Qy 1854 ATCTGACGCGGAGGCTGGGGGATTTTATTCAAAGATCTGTCGAGAGAAATGATCAT 1913  
Db 663 ATCGACGCTGACGAGATGGGGGCTTCATTAAGAACTCTGTCGATTAAGAAATGATCATCT 722  
Qy 1914 CTGAGATATCTGTGAGAGATTTATTTCTCAAGATGAGTCAGAGAGAGGAAATGTA 1973  
Db 723 TGGAGATGACTGTGTGATGATTTCTCAAGAGAGAGAGCAAAACGGGCAAAATTTA 782  
Qy 1974 TGAATAATACATGTGACGCTTTCTGTCACTTGAACAAATGATTTGTGTGTGATGCAAC 2033  
Db 783 TGACCGGCGCAAACTCTTCCTCTCTTGAATTAATGACAGTTTGTATGATGATGACAT 842  
Qy 2034 CGGCAAGGCTAACAAATTCGTTTTCGAATCATTCGGTAATCCAACTGCTATGCCAA 2093  
Db 843 TCGGAAGGGGATTAATTAAGTTGCAAAATCATCTTCACGCCCACTGCTATGACAA 902  
Qy 2094 AGTTATGATGTTAAACGCTGATCAACAGATAGGATTTTTCGCAAGAGAGCCATCCAGAC 2153  
Db 903 GGTGATGATGTGGCGGCTGACCATGCGGTGTGTATCTATGCAAGGAGCATATGTAAC 962  
Qy 2154 TGGCGAAGAGCTGTTTATGATTAAGATACAGATACAGCCAGGCTGATGCC 2200  
Db 963 TAGTGCCGAACCTCTCTATGATTAACCGGTATGAGACCGGACCAAGCCC 1009

## RESULT 5

US-09-699-266A-8  
; Sequence 8, Application US/09699266A  
; Patent No. 6559354  
; GENERAL INFORMATION:  
; APPLICANT: Ma, Hongchang  
; APPLICANT: Morakinyo, Layo O.  
; APPLICANT: Odell, Joan T.  
; APPLICANT: Orozco Jr., Emil M.  
; APPLICANT: Rafalski, J. Antoni  
; TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS  
; FILE REFERENCE: B01164 US NA  
; CURRENT APPLICATION NUMBER: US/09/699,266A  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: PCT/US99/08385  
; PRIOR FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/083,212  
; PRIOR FILING DATE: 1998-04-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 3148  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-699-266A-8

Query Match 6.2%; Score 138.2; DB 3; Length 3148;  
Best Local Similarity 62.7%; Pred. No. 4.9e-30;  
Matches 215; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 1841 TATTCCTGCGACCATCTGACGTGCGAGCGCTGGGGGATTTTATCAAGATCCTGTGACA 1900  
Db 2363 TCTTACTGGAAGATCAATGTCCTGCGCTGGGAGCATCTCTCAAGAAATAGTTAGCA 2422  
Qy 1901 AAAAAGATTCATCTCAGAAATCTGTGAGAGATTTATTTCTCAAGATGAAAGTCAGAGAA 1960  
Db 2423 AACATGAATACCTGTGTGTGATGACCTGGGAACTAATCTCACAAAGAAAGATTAAGC 2482  
Qy 1961 GAGGAAAGTATGATTAATATCATGTGACGCTTCTGTTCATTAAGTGAATGATTTTG 2020  
Db 2483 GTGGAAGATATATATATCTGTGAACTATCTTCTTCAACTGAAATGATGATG 2542  
Qy 2021 TGGTGAATGCAACCGGCAAGGGAACAAATTCGTTTTCGAATCATTCGTAATTCGA 2080  
Db 2543 TTTCTGACGATATGAGAAATGGGTGACAAAGCTGAATTTGCCAATGCTGACCGCA 2602

Qy 2081 ACTGTATGCAAAAGTTATGATGTTAAACGATGATCAAGATAGTATTTTGCCAGA 2140  
Db 2603 ATTGCTATGCCAAGTTATCATGATGATGATGATGATGATGATGATGATGATGATGATG 2662  
Qy 2141 GAGCATTCAGATGCGGCAAGAGCTGTTTATGATTAACAGATA 2183  
Db 2663 AAGATCTCTGCTGTGAGAGATTTATCTAGATTAACGCGTA 2705

## RESULT 6

US-09-177-249-1  
; Sequence 1, Application US/09177249  
; Patent No. 6229064  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Ohad, Nir  
; APPLICANT: Kiyosue, Tomohiro  
; APPLICANT: Yadegari, Ramin  
; APPLICANT: Margossian, Linda  
; APPLICANT: Harada, John  
; APPLICANT: Goldberg, Robert B.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
; FILE REFERENCE: 023070-086120US  
; CURRENT APPLICATION NUMBER: US/09/177,249  
; PRIOR FILING DATE: 1998-10-22  
; EARLIER APPLICATION NUMBER: US 09/071,838  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 324  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2136  
; TYPE: DNA  
; ORGANISM: Arabidopsis sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (43)..(2112)  
; OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1) CDNA  
US-09-177-249-1

Query Match 6.0%; Score 135.4; DB 3; Length 2136;  
Best Local Similarity 54.0%; Pred. No. 2.6e-29;  
Matches 333; Conservative 0; Mismatches 266; Indels 18; Gaps 2;

Qy 1600 TGGCTTGTGTGATGACAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGT 1659  
Db 1420 TGCCCTGTGTTAACTCAGAAATTTGCGAGAAATATTTGGGGTGTCTCAAGAGATTGC 1479  
Qy 1660 CAAAACGCTTTCGGGATGCGCGT---CAAAGCAGTGTGACACCAAGAGATGCCCG 1716  
Db 1480 AACATGCTTTGGAGATGTAATTTGTCAATTTGGCAATGACAAATGCAATGATGCTT 1539  
Qy 1717 TGCTACCTGCTGTGCGAGAGTGTGACCTGATCTGTCTTACTTGT----- 1764  
Db 1540 TGTTTTGTGCTTAATCGTGAATGATGATCAATCTTTGTGCGAGTGTCTCTTACGTGT 1599  
Qy 1765 ---GGAGCGCTGACCATTTGGACAGTAATAATGTCTCTGACAGATGCAATTCAG 1821  
Db 1600 GGAATGACACTCTTGTGAGACACAGTGCATAATCCATGCAAGAAATGCAATTCCTC 1659  
Qy 1822 CGGGGCTCCAAAAGATCTATATGTGTGACACATGTGACGTGCGAGGCTGGGGGATTTT 1881  
Db 1660 CTTCAACCAATTAATAAATATCTCAATTTGAAAGTGTGATGATGATGATGATGATG 1719  
Qy 1882 ATCAAGATCTGTGACAGAAATGAATTCATCTCAGATATCTGTGAGAGATTTTCT 1941  
Db 1720 ACATGGAATCTCTTAATAAAGATGATATCTCGAAGAAATATATCTGAGAACTATCAT 1779  
Qy 1942 CAAGATGAGCTGACAGAAAGGGAAGTGTATGATTAATATCATGTGACGCTTTCTGTT 2001  
Db 1780 CATGATGAAGCTAATGAGCGTGTGAGATTAAGATCGAATTTGTTCTTCTTACCTCTTT 1839

Qy	2002	AACCTGAAACAATGATTTTGTGTGGATCAACCCGCAAGGGGTAAACAATAATCGTTTTCA	2064
Db	1840	ACCTTGATTAATCAGCTGGAATTCATGCTGCCTGCTAAGAACGATTCAAATTTCTC	1895
Qy	2062	AATCATTCGGTAAATCCAACCTGCTATGCAAAAAGTTATGATGTTAACGGTATCACAGG	2121
Db	1900	AATCACTCAGCAAGACTTAATCTGTAACGCCAAGGTATGATGTGTAGAAGAGATCAGAGG	1958
Qy	2122	ATAGGATATTTTTCGCAAGAGACCATCCAGACTGGCGAAGAGCTGTTTTGATTAACGA	2181
Db	1960	ATTGGTCTAATTTTCGGAAGAGCAATCGAAGAAAGGTAGAGAGCTTTCTTGACTAATGTC	2015
Qy	2182	TACAGCCAGGCTGATGC	2198
Db	2020	TATGACCAAGAACTGC	2036

RESULT 7  
US-09-81

Sequence 1 Application US/09812283  
Patent No. 6828477  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Onad, Nlr  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadegari, Ramln  
APPLICANT: Margossian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/812,283  
CURRENT FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 09/177,249  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: US 09/071,838  
PRIOR FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 1  
LENGTH: 2136  
TYPE: DNA  
ORGANISM: *Arabidopsis* sp.  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (43)..(2112)  
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1) cDNA  
US-09-812-283-1

Query Match	Similarity	6.0%	Score 135.4	DB 3	Length 2136
Best Local	Similarity	54.0%	Pred. No. 2.6e-29		
Matches	333	Conservative	0	Mismatches 266	Indels 18
				Gaps	2

  

Qy	1600	TGCCCTGTGTGATGACAAATATTTGTGTGAAAATTTCATATGATTCAGAGTGT	1655
Db	1420	TGCCCTTGTTTAATCAAGAAATTTGTGTGCGAATAATATGGGGGTCTCAAGATATGC	1475
Qy	1660	CAAAACCGCTTCCGGGATGCCGCTG---CAAGACAGTGCACACCAAGAGTGCCTG	1715
Db	1480	AACATCGCTTTGGAGAGATGTATTTGTGCATTTGGCAATGTGCATGACAAATCGAATGCTCT	1535
Qy	1717	TGCTACGTGGGCTGTCCGAGAGTGTACCCGTGACTGTCTTACTTGT-----	1766
Db	1540	TGTTTGTCTGCTAATGTGTAAATGCAATCCAGATCTTTGTGCGAGTTGTCTCTTAGCTGT	1599
Qy	1765	---GGAGCGCTGACCATTTGGGACAGTAAATGTGTCTCTGCAGAGTACAGTATTCAG	1821
Db	1600	GGAAGATGGACATCTTGTGTGAGACACAGTGCAAATCCATATGAAAGACATGTAAATTCCTC	1655
Qy	1822	CGGGGCTCCAAAAGCATTTATTTGCTGGACCATTTGACGTGGCAGGCTGGGGGATTTTT	1881

Db	1660	CTTGAAACCAATAAAAAAGATTCTCATTTGGAAAGCTCATGTTTCATGTAGTGGGGTGCAATT	1719
Qy	1882	ATCAAAGATCCTGTGCAGAAAAATGAATTCACTCAGAAATACGTGGAGATTAATTCT	1941
Db	1720	ACATGGACCTCTCTTAAAAAGAAAGATATCTCGGAAAAATATACGTGAGAACATCACT	1779
Qy	1942	CAAAATGAGCTGCAGAAAGGGGAAAGTGATGATTAATACATGTGAGCTTCTGTC	2001
Db	1780	CATATGAAAGCTAATGAGCGTGGAGAAATGAAAGATCGATTGGTTCTTCTCACTCTTT	1839
Qy	2002	AACTTGAACAATGATTTTGTGTGATGCAACCGCAAGGTAACAAAATTGTTTTGCA	2061
Db	1840	ACCTTGAATGATGAGCTGGAATGATGCTGCCGTAAAGGAAACGATTCAAAATTCTC	1899
Qy	2062	AATCATTCGGTAAATCCAAACTGCTATGCAGAAAGTATGATGTTAAACGTGATCAAGG	2121
Db	1900	AATCACTCAGCAAGACCTTACTGCTACGCCAAGTGTGATGATTTGAGAGGAGATCAGAG	1959
Qy	2122	ATTAGTATTTTGGCAGAGAGCCATCCAGACTGGCGAAGAGCTGTTTTTGATTACGA	2181
Db	1960	ATTGCTCATTTTGGGAGAGAGCAATCGAAGAGGTAGAGGCTTTTCTTGACTACTGC	2019
Qy	2182	TACAGCCAGCGCTGATGC	2198
Db	2020	TATGACCAAGAACTGC	2036

**RESULT 8**  
**INC-09-07**

Sequence 1, Application US/09071838A  
Patent No. 7029917  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
Ohad, Nir  
Kiyosue, Tomohiro  
Vadegari, Ramin  
Margossian, Linda  
Harada, John  
Goldberg, Robert B.  
TITLE OF INVENTION: Nucleic Acids That Control Seed and  
Fruit Development in Plants  
NUMBER OF SEQUENCES: 324  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,838A  
FILING DATE: 01-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Baetian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-086100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2136 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:



```

FILE REFERENCE: 023070-116710US
CURRENT APPLICATION NUMBER: US/10/176,884
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/300,506
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 2799
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: MEA-15G-ESAI
OTHER INFORMATION: chimeric HAT nucleic acid construct encoded by
OTHER INFORMATION: transgene plasmid pCR304-5UNS MEA-15G-ESAI
US-10-176-884-16

```

Query Match	6.0%;	Score 135.4;	DB 3;	Length 2759;
Best Local Similarity	54.0%;	Pred. No. 3.1e-29;		
Matches 333;	Conservative 0;	Mismatches 266;	Indels 18;	Gaps 2

Qy	1600	TGCCCTTGATGACAAAAATTTTGTGAAAGTTTGTGCATGTATGCAAGT	165
Db	1378	TGCGCTTTGTTACTCAGAAAATTGCTGCGAATAATATGGGGTGTCTCAAGATTC	143
Qy	1660	CAAAACCGCTTCCGGGATGCCGTG---CAAGAACAAGTGAACCAAGACGTGCCG	1716
Db	1438	AACATATCGCTTTGAGAGATGTAAATGTGCATTTGGCCAAATGCAAAATCGAATGTCT	1497
Qy	1717	TGCTACCTGGCTGTGCCGAGTGTACCTGTACCTGTCTTACTTGT-----	1766
Db	1498	TGTTTGTGCTAAATGTGAATGCATCCAGATCTTTGTGCGAGTTGTCTCTTAGCTGT	1557
Qy	1765	---GGAGCGGTCACATTTGGGACAGTAAATAATGTCTCTGCAGAACTGCAGTATTCAG	1822
Db	1558	GGAGATGGCACTCTTGATGAGACACAGTGCAAATCCATATGCAAGAACATGCATTTCTC	161
Qy	1822	CGGGGCTCCAAAAGCATCTATTTGCTGGCCATCTGACGTGCGAGGCTGGGGGAAATTTT	188
Db	1618	CTTCAAAACCAATAAAAAGATTCCTCATTTGGAAAGTCTGATGTTCATGAAATGGGGTGCATTT	1677
Qy	1882	ATCAAGAATCCGTGCAGAAAATGAATTATCTAGAAATACGTGGAGAGATTTATTTCT	1941
Db	1678	ACATGGGACTCTTTAAABAAGATGATGTCTTCGGAGATTAATCTGGAGAACTGATCACT	173
Qy	1942	CAGAATGAAGCTGACAGAGAGGGAAAGTATGATTAATACATGTGACGCTTTCGTTTC	2001
Db	1728	CATGATGAAGCTAATGAGCGTGGAGAAATAGAAGATCGGATTTGGTTCTTCTCACTCTTT	179
Qy	2002	AACCTGGAACAATGATTTTGTGTGTGATGCAACCCGCAAGGTGAACAAAATTCGTTTGA	206
Db	1798	ACCTTGAATGATCACTCGAATTCGAATGCATGTCCCGTAAAGGAACGAGTTCAAAATTTCTC	1857
Qy	2062	AATCATTCGGTAAATCCAAACTGTCTATGCAAAAGTATGATGTTTAAACGTGATCAAGG	2122
Db	1858	AATCATCTGAGCAAGACCTAATCTGCTACGCGCAAGTTGATGATTTGTGAGGAAGATCAAGG	191
Qy	2122	ATAGGATATTTTGCAGAGAGACCATCCAGACTGGCGAAAGAGCTGTTTTTGAATTAACA	218
Db	1918	ATTGTCTATTTTGCAGAGAGACCAATCGAAGAAGGTGAGAGACTTTTCTTCACATACTGC	197
Qy	2182	TACACCCAGCGTGATGC 2198	
Db	1978	TATGAGCCAGAACTATGC 1994	

RESULT 11  
US-09-640-211A-349  
; Sequence 349, Application US/09640211A  
; Patent No. 683346  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Marion  
; APPLICANT: Shenk, Michael A.

```

; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CU
; CURRENT APPLICATION NUMBER: US/09/640.211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 349
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Pinus radiata
; US-09-640-211A-349

```

Query Match	5.3%	Score 119.2;	DB 3;	Length 260;
Best Local Similarity	68.0%	Pred. No. 4.5e-25;		
Matches 166;	Conservative 0;	Mismatches 78;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 12
US-09-640-211A-591
; Sequence 591, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 591
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-640-211A-591

```

Query Match	5.3%	Score 119.2	DB 3	Length 260
Best Local Similarity	68.0%	Pred. No. 4.5e-25		
Matches 166	0	Mismatches 76	Indels 0	Gaps 0

Qy	1918	GAATACGTGGAGATTAATTTCCAAATGAGGTGCACAAAGGGGAAAGTATGAT	1977
Db	17	GAGTACTGAGAGCTTGAATTTTCACATGGGAGAGCTGTAAAGCAGGAAAAGTTATGAT	76
Qy	1978	AAATACATGTGACGTTCTTGTCACATTGAAACATGATTTTGTGTGATGCAACCCG	203
Db	77	CGAGAGACCTCTCTCTCTTTTCAACTTGAAACATGACGATGTCGTTGATGACATACCG	136



QY 2038 AAGGTAAACAATTCGTTTGCATATCATTCGTAAATCCAAATCTGCTAATGCAAAAGT 2097  
DB 137 AAGGGGAGTAAGTAAATTCATTAATCAATCCAAATCTGCTAATGCAAAAGT 196  
QY 2098 ATGATGTGTAACGTGATCAGATAGATATTTTGGCCAAAGAGCCATCCAGACTGCGC 2157  
DB 197 ATTAAGTTCGCTGATGATCATAGATGGTATTTTGCAGAGAGAGATTCGACCCGCT 256  
QY 2158 GAAG 2161  
DB 257 GAGG 260

## RESULT 13

US-09-061-769A-4  
; Sequence 4, Application US/09061769A  
; Patent No. 6239327  
; GENERAL INFORMATION:  
; APPLICANT: Cold Spring Harbor Labs  
; TITLE OF INVENTION: No. 6239327e1 Seed Specific Polycarb Group Gene and Methods of  
; TITLE OF INVENTION: use for Same  
; FILE REFERENCE: mea genes  
; CURRENT APPLICATION NUMBER: US/09/061,769A  
; PRIOR FILING DATE: 1998-04-16  
; PRIOR APPLICATION NUMBER: 09/061,769  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 2322  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: MEA2  
; NAME/KEY: variation  
; LOCATION: (1765)..(1763)  
US-09-061-769A-4

Query Match 5.3%; Score 118.4; DB 3; Length 2322;  
Best Local Similarity 53.4%; Pred. No. 3.2e-24;  
Matches 333; Conservative 0; Mismatches 266; Indels 25; Gaps 3;

QY 1600 TCCCTTGTGATGACACAAATTTTGTGAAGTTTGCATGATGTTCAAGT 1659  
DB 1391 TGCCCTGTGTAATCAAGAAATTTGTCGAGAAATTTGCGGTGCTCAAGAGATTGC 1450  
QY 1660 CAAAACCGCTTCCGGGATGCGCGT---CAAGACAGTGCACACCAAGAGTCCCG 1716  
DB 1451 AACATCGCTTGGAGATGTAATTTGCAATTTGCCAATGCACAAATGCAATGCTCT 1510  
QY 1717 TGCTACCTGGCTGTCGAGAGTGAACCTGACCTGCTCTTACTTGT----- 1764  
DB 1511 TGTTCGCTGTAATCGTAATGCGATCAATCTTTTGGAGTTGTCCTTAGCTGT 1570  
QY 1765 ---GAGCGCTGACCAATTTGGACAGTAAATGTCCTGCAAGAACTGCAATTTAG 1821  
DB 1571 GGAATGACACTTGTGAGACACCAAGTCAATCAATGCACAAATGCAATTTCTC 1630  
QY 1822 CGGGGCTCCAAAAGCATCTATTGCTGCAACATCTGACGAGCGCGGGGATTTT 1881  
DB 1631 CTTCACCAATTAATAAGATTTCTCATGAAATGTCATGATGATGAGGAGTCAATTT 1690  
QY 1882 ATCAAGATCTGTCAGAGAAATGATTCATCTGAGATATCTGAGAGATTTATTC- 1940  
DB 1691 ACATGGAGCTCTTTAAAAGATGATATCTCGAGAAATATACCTGAGAACTGATCACT 1750  
QY 1941 -----TCAAGTGAAGTGCAGAGAGGAAAGTGTATGATTAATCAATGCACTT 1994  
DB 1751 CATGATCATGATGAGCTAATGAGCGTGGGAGATGAGATCGAATTTGTTCTTCTTA 1810  
QY 1995 TCTGTTCACTTGAACATGATTTTGTGTGATGCAACCGCGAGAGGTAACAAATTCG 2054

DB 1811 CCTTTTACCTTGAATGATCAGCTGAAATGATGCTGCCGCTAAAGAAACGAGTTCA 1870  
QY 2055 TTTTGCAATCATTTGCGTAATTCCAAATCTGCTAATGCAAAAGTATGATGTTAAGCTGA 2114  
DB 1871 ATTTCTCAATCACTCAGAGAAAGCTTACCTGCTAAGCAAGTTGATGATGAGAGAGA 1930  
QY 2115 TCACAGATGATGATTTTGGCCAAAGAGCCATCCAGACTGAGCGAGAGAGCTGTTTGA 2174  
DB 1931 TCAGAGATTTGCTCTATTGTCGAGAGAGCAATGCAGAAAGTGAAGAGCTTTCTTGA 1990  
QY 2175 TTACAGATTAACAGCCAGCTGATGC 2198  
DB 1991 CTACTGCTATGAGCAGAAATGCT 2014

## RESULT 14

US-09-061-769A-3  
; Sequence 3, Application US/09061769A  
; Patent No. 6239327  
; GENERAL INFORMATION:  
; APPLICANT: Cold Spring Harbor Labs  
; TITLE OF INVENTION: No. 6239327e1 Seed Specific Polycarb Group Gene and Methods of  
; TITLE OF INVENTION: use for Same  
; FILE REFERENCE: mea genes  
; CURRENT APPLICATION NUMBER: US/09/061,769A  
; PRIOR FILING DATE: 1998-04-16  
; PRIOR APPLICATION NUMBER: 09/061,769  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2327  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: MEA1  
; NAME/KEY: variation  
; LOCATION: (1760)..(1761)  
US-09-061-769A-3

Query Match 5.1%; Score 115; DB 3; Length 2327;  
Best Local Similarity 53.1%; Pred. No. 3.3e-23;  
Matches 334; Conservative 0; Mismatches 265; Indels 30; Gaps 3;

QY 1600 TCCCTTGTGATGACACAAATTTTGTGAAGTTTGCATGATGTTCAAGT 1659  
DB 1391 TGCCCTGTGTAATCAAGAAATTTGTCGAGAAATTTGCGGTGCTCAAGAGATTGC 1450  
QY 1660 CAAAACCGCTTCCGGGATGCGCGT---CAAGACAGTGCACACCAAGAGTCCCG 1716  
DB 1451 AACATCGCTTGGAGATGTAATTTGCAATTTGCCAATGCACAAATGCAATGCTCT 1510  
QY 1717 TGCTACCTGGCTGTCGAGAGTGAACCTGACCTGCTCTTACTTGT----- 1764  
DB 1511 TGTTCGCTGTAATCGTAATGCGATCAATCTTTTGGAGTTGTCCTTAGCTGT 1570  
QY 1765 ---GAGCGCTGACCAATTTGGACAGTAAATGTCCTGCAAGAACTGCAATTTAG 1821  
DB 1571 GGAATGACACTTGTGAGACACCAAGTCAATCAATGCACAAATGCAATTTCTC 1630  
QY 1822 CGGGGCTCCAAAAGCATCTATTGCTGCAACATCTGACGAGCGCGGGGATTTT 1881  
DB 1631 CTTCACCAATTAATAAGATTTCTCATGAAATGTCATGATGATGAGGAGTCAATTT 1690  
QY 1882 ATCAAGATCTGTCAGAGAAATGATTCATCTGAGATATCTGAGAGATTTATTC- 1937  
DB 1691 ACATGGAGCTCTTTAAAAGATGATATCTCGAGAAATATACCTGAGAACTGATCACT 1750  
QY 1938 -----TTCAGATGAAGTGCAGAGAGGAAAGTGTATGATTAATCAATGCTG 1989  
DB 1751 CACTCATGATCTCATGATGAGCTAATGAGCGTGGGAGATGAGATCGAATTTGTTCT 1810  
QY 1990 AGCTTTCGTTCACTTGAACATGATTTTGTGTGATGCAACCGCGAGAGGTAACAA 2049

Accession	Sequence	Position
Db	TCCTACCTCTTACCTGGAATGATCGACTCGAAATCGATGCTCGCCGTAAGGAAGACAG	1870
Oy	ATTGCTTTGGCAATCATTCGGTAAATCCAAACTGCTATGCAAAAGTTATGATGTTAAC	2109
Db	TTCAAATTTCCAAATCACTCAGCAAGACTTAACTGCTACGCCAAGTTGATGTTGAGA	1930
Oy	GGTATCACAGATAGGTATTTTGGCCAAAGAGACCATCCAGACTGGCGAAGACTGTTT	2169
Db	GGAGTTCAGAGGATTTGTCCTATTTTGGGAGAGAGCATCGAAGAGGTGAGGACTTTTC	1990
Oy	TTTGATTAACAGATACAGCCAGGCTGATGC	2198
Db	TTTCAGTACTGCTATGACACGAACATTC	1991

```

RESULT 15
US-09-270-767-765/C
; Sequence 765, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: file Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 765
; LENGTH: 659
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-765

```

Query Match	3.1%	Score 70.4	DB 3	Length 659
Beet Local Similarity	52.0%	Pred. No. 3e-10		
Matches 186	Conservative	0	Mismatches 166	Indels 6
			Gaps	1
Qy	1831	AAAAAGCATCTATTGCTGTGGCAACATCTGACCTGGGCGAGATTTTATCCAAAGAT	1890	
Db	502	AAAGAAACGCTTAAAGTTTGCCAAATCTCGATACATGATTTGGGATTTATTGGCAATGGAG	443	
Qy	1891	CCTGTGCAGAAAAATGAAATTGATCTCAGAAATACGTGTGGAGAGATTATTCTCCAAAGATGAA	1950	
Db	442	CCCATAGCTGCAGAGCAAAATGGTTATTGAAATATGTTGGTCAGATGATTCGACCTGATGTT	383	
Qy	1951	GCTGCAGAAAGGAGGAAAGTGTATGATTAATATACATGTGCAG-----CTTCTGTCTCAAC	2004	
Db	382	GCTGATCTTAAGAGAAACAAAGTATATACAAATGGGATTTGGAGTCTCTAATTTTGTTCGA	323	
Qy	2005	TTGAACATATGATTTTGTGTGGATGCAACCCGCAAGGGTAAACAAATTCGTTTGGCAAT	2064	
Db	322	ATTGACATGGAACATATATATTGACGCACTAAATGTGAAACTTAGTCGGTTTATTAAT	263	
Qy	2065	CATTGCGTAAATCCAAACTGCTATGCAAAAGTTATGATGTTAAACGATGATCACAAGATA	2124	
Db	262	CACAGTGCATCCGAATGTCTACCGCTAAGGTTATATCTAATGATCTGAAAAGAGATA	203	
Qy	2125	GGTATTTTGGCCAAAGAGCCATCCAGACTGCGGAAAGCGTGTTTTGTATTAACGAT	2182	
Db	202	GTAATATATTCAAAGCAACCCATGGTATTCACACGGAGAAATTAACCTTACGACTATTAAT	145	

Search completed: August 13, 2006, 16:47:14  
Job time : 449 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2006, 11:51:22 ; Search time 2816 Seconds

(without alignments)  
9778.619 Million cell updates/sec

Title: US-10-773-302-1\_COPY\_90\_2330

Perfect score: 2241

Sequence: 1 atggccgcgactgggaagaa.....gagaatggaatcccttga 2241

Scoring table: IDENTITY\_NUC

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## ALIGNMENTS

18	828.8	37.0	4185	10	US-10-764-420-1602	Sequence 1602, Ap
19	713.6	31.8	2522	7	US-10-115-482-55	Sequence 55, Appl
20	594	26.5	807	8	US-10-424-599-76875	Sequence 76875, A
21	582.2	26.0	714	9	US-10-333-872A-67	Sequence 67, Appl
22	555	24.8	555	3	US-09-796-692-5700	Sequence 5700, Ap
23	555	24.8	555	6	US-10-040-862-5700	Sequence 5700, Ap
24	555	24.8	555	7	US-10-057-475B-5700	Sequence 5700, Ap
25	555	24.8	555	7	US-10-154-884B-5700	Sequence 5700, Ap
26	555	24.8	555	9	US-10-764-324-5700	Sequence 5700, Ap
27	514	22.9	548	3	US-09-796-692-5422	Sequence 5422, Ap
28	514	22.9	548	6	US-10-040-862-5422	Sequence 5422, Ap
29	514	22.9	548	7	US-10-057-475B-5422	Sequence 5422, Ap
30	514	22.9	548	7	US-10-154-884B-5422	Sequence 5422, Ap
31	514	22.9	548	3	US-10-764-324-5422	Sequence 5422, Ap
32	428	19.1	485	3	US-09-918-995-21967	Sequence 21967, A
33	428	19.1	2448	13	US-11-097-143-14072	Sequence 14072, A
34	351.6	15.7	5289	13	US-11-097-143-14071	Sequence 14071, A
35	259.6	11.6	565	8	US-10-424-599-55297	Sequence 55297, A
36	253	11.3	253	6	US-10-060-036-2840	Sequence 2840, Ap
37	252	11.2	417	3	US-09-864-761-4532	Sequence 4532, Ap
38	242.4	10.8	430	3	US-09-864-761-14986	Sequence 14986, A
39	241.4	10.8	295	16	US-11-021-492-277	Sequence 277, App
40	241	10.8	308	3	US-09-864-761-31520	Sequence 31520, A
41	218.8	9.8	489	10	US-10-773-302-5	Sequence 5, Appl1
42	215.6	9.6	230	3	US-09-864-761-21279	Sequence 21279, A
43	208.6	9.3	4846	13	US-11-097-143-17719	Sequence 17719, A
44	176.8	7.9	448	8	US-10-430-201-207	Sequence 207, App
45	176.8	7.9	448	8	US-10-430-201-208	Sequence 208, App

Database : Published Applications NA Main:

1: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
3: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
4: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
5: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
6: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
7: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
8: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
9: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
10: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
11: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
12: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
13: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
14: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq:\*  
15: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq:\*  
16: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2241	100.0	2600	3	US-09-954-456-1157
2	2241	100.0	2600	10	US-10-843-641A-4184
3	2241	100.0	2600	10	US-10-773-302-1
4	2241	100.0	2711	9	US-10-723-860-5458
5	2236.2	99.8	2558	6	US-10-153-668-231
6	2236.2	99.8	2576	7	US-10-172-118-1043
7	2236.2	99.8	2576	7	US-10-295-027-709
8	2236.2	99.8	2576	8	US-10-342-887-1043
9	2236.2	99.8	2576	10	US-10-848-755A-129
10	2236.2	99.8	2576	10	US-10-504-173-78
11	2236.2	99.8	2576	12	US-10-960-414-396
12	2229	99.5	2512	7	US-10-210-120-95
13	2229	99.5	2512	7	US-10-909-035-95
14	1993.8	89.0	2476	7	US-10-104-047-1192
15	1993.8	89.0	2476	16	US-11-072-512-1192
16	1700	75.9	1960	6	US-10-153-668-353
17	840.4	37.5	2594	10	US-10-504-173-77

RESULT 1

US-09-954-456-1157

Sequence 1157, Application US/09954456

Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Car

FILE REFERENCE: 689290-76

CURRENT APPLICATION NUMBER: US/09/954, 456

CURRENT FILING DATE: 2001-09-18

PRIOR APPLICATION NUMBER: US/60/233, 617

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: US/60/234, 052

PRIOR FILING DATE: 2000-09-20

PRIOR APPLICATION NUMBER: US/60/234, 923

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235, 134

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235, 637

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235, 638

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235, 711

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235, 720

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235, 840

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235, 863

PRIOR FILING DATE: 2000-09-27

NUMBER OF SEQ ID NOS: 2276

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1157

LENGTH: 2600

TYPE: DNA

ORGANISM: Homo sapiens

US-09-954-456-1157

Query Match 100.0%; Score 2241; DB 3; Length 2600;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGGCCGAGCTGGGAAGAATCTGAGAAGGACAGGTTTGTGGCGGAACGGTGTAAA 60
Db 90 ATGGCCGAGCTGGGAAGAATCTGAGAAGGACAGGTTTGTGGCGGAACGGTGTAAA 149
OY 61 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTCAAGACGATGATGAAGTAAAGT 120
Db 150 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTCAAGACGATGATGAAGTAAAGT 209
OY 121 ATGTTAGTCCAAATCGTCAAGAAAATTTGGAAAAGAACGAAATCTTAAACCAAGAAATG 180
Db 210 ATGTTAGTCCAAATCGTCAAGAAAATTTGGAAAAGAACGAAATCTTAAACCAAGAAATG 269
OY 181 AAACGCGAAGGATPACGCTGTGACATCCTGACTTGTGAGCTCATTTGCGGGGACT 240
Db 270 AAACGCGAAGGATPACGCTGTGACATCCTGACTTGTGAGCTCATTTGCGGGGACT 329
OY 241 AGGAGTGTGCGTGAACGATGACTTGGATTTTCCAAACAAGTCATCCATTAAAGACT 300
Db 330 AGGAGTGTGCGTGAACGATGACTTGGATTTTCCAAACAAGTCATCCATTAAAGACT 389
OY 301 CTGAATGCAATGCTTCAAGTACCCATTAATGTAATCTTGGTCTCCCTTACAGAGATTTT 360
Db 390 CTGAATGCAATGCTTCAAGTACCCATTAATGTAATCTTGGTCTCCCTTACAGAGATTTT 449
OY 361 ATGTGGAAGATGAACGTTTTCATTAACATTCCTTATATGGGAGATGAAGTTTAGAT 420
Db 450 ATGTGGAAGATGAACGTTTTCATTAACATTCCTTATATGGGAGATGAAGTTTAGAT 509
OY 421 CAGAGTGTACTTTCATTTGAAGAACTAATTAATAAATTTATGATGGAAAGTACACGGGAT 480
Db 510 CAGAGTGTACTTTCATTTGAAGAACTAATTAATAAATTTATGATGGAAAGTACACGGGAT 569
OY 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGTAATGCCCTTGTGTA 540
Db 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGTAATGCCCTTGTGTA 629
OY 541 TATATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 630 TATATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
OY 601 AAAGATCTGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 660
Db 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 749
OY 661 GATTAATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAACGACGAAAGACTA 720
Db 750 GATTAATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAACGACGAAAGACTA 809
OY 721 AAAGAAAATATTAAGAACTGACGGAACGAGCTCCAGGGGCACTTCTCTGTAATGT 780
Db 810 AAAGAAAATATTAAGAACTGACGGAACGAGCTCCAGGGGCACTTCTCTGTAATGT 869
OY 781 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTTACACTCC 840
Db 870 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTTACACTCC 929
OY 841 TTTCAATGCGTTTCTGTAGCGGATGTTTTAAATATGACTGCTTCTCAATCTTTTCAT 900
Db 930 TTTCAATGCGTTTCTGTAGCGGATGTTTTAAATATGACTGCTTCTCAATCTTTTCAT 989
OY 901 GCAACACCAACACTTAATTAAGCGGAAGAACAGAAACGCTTAAGACAAACAACTTGT 960
Db 990 GCAACACCAACACTTAATTAAGCGGAAGAACAGAAACGCTTAAGACAAACAACTTGT 1049
OY 961 GGAACCAAGTGTATACAGATTTTGGAGGAGCAAAAGGTTTGTGCTGCTTCAACCGCT 1020
Db 1050 GGAACCAAGTGTATACAGATTTTGGAGGAGCAAAAGGTTTGTGCTGCTTCAACCGCT 1109
OY 1021 GAGCGATTAAGACCCCAACAAACGTCAGAGGCGCGAAGAGAGCGCTTCCCAAT 1080
```

```
Db 1110 GAGCGATTAAGACCCCAACAAACGTCAGAGGCGCGAAGAGAGCGCTTCCCAAT 1169
OY 1081 AACAGTACAGGCGCCAGACACCCCAACCAATTAATGTGCTGTAATCAAGATACAGACGT 1140
Db 1170 AACAGTACAGGCGCCAGACACCCCAACCAATTAATGTGCTGTAATCAAGATACAGACGT 1229
OY 1141 GATAGGAGACAGGGACTGAAGCGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1200
Db 1230 GATAGGAGACAGGGACTGAAGCGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1289
OY 1201 AAAGATGAATCTTGACTGCTCTGAGCAAAATTCGCTGTCAAAACCAATTAAGATG 1260
Db 1290 AAAGATGAATCTTGACTGCTCTGAGCAAAATTCGCTGTCAAAACCAATTAAGATG 1349
OY 1261 AAGCCAATATTAAGACCTCTGAGAAATGTGAGTGAAGTGTGTAAGCTCAATGTTT 1320
Db 1350 AAGCCAATATTAAGACCTCTGAGAAATGTGAGTGAAGTGTGTAAGCTCAATGTTT 1409
OY 1321 AGAGTCTCATTTGCACTTACTATGACAAATTTCTGTGCCATTTGTAGGTTAATGGGACC 1380
Db 1410 AGAGTCTCATTTGCACTTACTATGACAAATTTCTGTGCCATTTGTAGGTTAATGGGACC 1469
OY 1381 AAAACATGTAGACAGGTGATGAGTTAAGTCAAGAAATCTAGCATCATAGCTCAAGCT 1440
Db 1470 AAAACATGTAGACAGGTGATGAGTTAAGTCAAGAAATCTAGCATCATAGCTCAAGCT 1529
OY 1441 CCCGCTGAGATGTGATATCTCTCAAGGAAAAGAAAGAAACAACGGTGTGGGCT 1500
Db 1530 CCCGCTGAGATGTGATATCTCTCAAGGAAAAGAAAGAAACAACGGTGTGGGCT 1589
OY 1501 GCACATGCGAAGAAAGATPACAGCTGAAGAAAGACGGCTCTCTTAACATGTTTACACTAT 1560
Db 1590 GCACATGCGAAGAAAGATPACAGCTGAAGAAAGACGGCTCTCTTAACATGTTTACACTAT 1649
OY 1561 CAACCTGTGATCATCAACGGGACCTTGTGACAGTTCGTGCCCTGTGTGATGACAA 1620
Db 1650 CAACCTGTGATCATCAACGGGACCTTGTGACAGTTCGTGCCCTGTGTGATGACAA 1709
OY 1621 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTGCAAAACGGCTTCCGGAGTGC 1680
Db 1710 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTGCAAAACGGCTTCCGGAGTGC 1769
OY 1681 CGCTGCAAGACAGCTGCAACACAGACAGTCCCGGTCTACTCGTGTCTCCAGAGACT 1740
Db 1770 CGCTGCAAGACAGCTGCAACACAGACAGTCCCGGTCTACTCGTGTCTCCAGAGACT 1829
OY 1741 GACCTGACCTCTGTCTTAATCTGTGAGCGGCTGACCAATTTGGGACAGTAAATGTGTCC 1800
Db 1830 GACCTGACCTCTGTCTTAATCTGTGAGCGGCTGACCAATTTGGGACAGTAAATGTGTCC 1889
OY 1801 TGCAAGAACTGCAAGTATTCAGCGGGGCTCCAAAAGCAATCTATTGTGGCAACATCTGAC 1860
Db 1890 TGCAAGAACTGCAAGTATTCAGCGGGGCTCCAAAAGCAATCTATTGTGGCAACATCTGAC 1949
OY 1861 GTGGCAGGCTGGGGGATTTTATCAAGAGATCTGTGCAAGAAATGAAATTCATCTCGAA 1920
Db 1950 GTGGCAGGCTGGGGGATTTTATCAAGAGATCTGTGCAAGAAATGAAATTCATCTCGAA 2009
OY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGATGATA 1980
Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGATGATA 2069
OY 1981 TACATGTGACAGCTTCTGTCTCAACTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2040
Db 2070 TACATGTGACAGCTTCTGTCTCAACTTGAACATGATTTTGTGTGATGCAACCCGCAAG 2129
OY 2041 GGTAAACAAATTTGTTTTCGTAATCATTTCCGTTAATCCAACTGCTATGCAAAAGTTATG 2100
Db 2130 GGTAAACAAATTTGTTTTCGTAATCATTTCCGTTAATCCAACTGCTATGCAAAAGTTATG 2169
OY 2101 ATGTTTAAACGGTATPACAGATAGGTATTTTGTGCAAGAGAGCATCTGCGGAA 2160
Db 2190 ATGTTTAAACGGTATPACAGATAGGTATTTTGTGCAAGAGAGCATCTGCGGAA 2249
```

QY 2161 GAGCTGTTTTGATTACAGATACAGCGGCTGATGCCCTGAAGTATGCGCATCGAA 2220  
|  
DB 2250 GAGCTGTTTTGATTACAGATACAGCGGCTGATGCCCTGAAGTATGCGCATCGAA 2309  
|  
QY 2221 AGAGAAATGGAATCCCTTGA 2241  
|  
DB 2310 AGAGAAATGGAATCCCTTGA 2330  
|

## RESULT 2

US-10-843-641A-4184  
; Sequence 4184, Application US/10843641A  
; Publication No. US20050064454A1  
; GENERAL INFORMATION:  
; APPLICANT: Avalon Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; FILE REFERENCE: 689290-189  
; CURRENT APPLICATION NUMBER: US/10/843,641A  
; CURRENT FILING DATE: 2004-05-12  
; PRIOR APPLICATION NUMBER: US/09/873,367  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US/09/954,531  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,436  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,832  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/964,824  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/09/967,768  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US/09/968,007  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,347  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,708  
; PRIOR FILING DATE: 2001-10-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 8447  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4184  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-843-641A-4184

Query Match 100.0%; Score 2241; DB 10; Length 2600;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCAGACTGGGAAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAAGCTCTAAA 60  
DB 90 ATGGGCGCAGACTGGGAAGAAATCTGAGAAAGGACCAAGTTGTTGGCGGAAGCTCTAAA 149  
|  
QY 61 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAAGTAAAGAGT 120  
DB 150 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAAGTAAAGAGT 209  
|  
QY 121 ATGTTAGTTCCATGCTGAGAAAATTTTGAAGAAGCGAAATCTTAAACCAAGATGG 180  
DB 210 ATGTTAGTTCCATGCTGAGAAAATTTTGAAGAAGCGAAATCTTAAACCAAGATGG 269  
|  
QY 181 AAACGCGAAGAGTACAGCTGCTGACATCTGAGCTTCTGAGCTCATTTGGCGGAGCT 240  
DB 270 AAACGCGAAGAGTACAGCTGCTGACATCTGAGCTTCTGAGCTCATTTGGCGGAGCT 329  
|  
QY 241 AGGAGTGTTCGTGACCAAGTGAATTGTTTCCAAACAAGTCAATCCCAATTAAAGACT 300  
DB 330 AGGAGTGTTCGTGACCAAGTGAATTGTTTCCAAACAAGTCAATCCCAATTAAAGACT 389  
|

QY 301 CTGAATGAGTTGCTTCAGTACCCATATATGATTTCTTGCTGCCCTACAGCAAAATTT 360  
DB 390 CTGAATGAGTTGCTTCAGTACCCATATATGATTTCTTGCTGCCCTACAGCAAAATTT 449  
|  
QY 361 ATGTGGAAGATGAAACTGTTTACATTAACATTCCTTATATGAGATGAAGTTTACAT 420  
DB 450 ATGTGGAAGATGAAACTGTTTACATTAACATTCCTTATATGAGATGAAGTTTACAT 509  
|  
QY 421 CAGATGCTACTTTCATTTGAAGAACTAATTAATAAATTAATGATGGAAAGTACCGGAT 480  
DB 510 CAGATGCTACTTTCATTTGAAGAACTAATTAATAAATTAATGATGGAAAGTACCGGAT 569  
|  
QY 481 AGAGAAATGGGTTTATTAATGATGAATTTTGTGGAGTTGATGAATGCCCTTGCTCA 540  
DB 570 AGAGAAATGGGTTTATTAATGATGAATTTTGTGGAGTTGATGAATGCCCTTGCTCA 629  
|  
QY 541 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 630 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
|  
QY 601 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660  
DB 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749  
|  
QY 661 GATTAATTTTGTGAAGCAATTTCTGCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 720  
DB 750 GATTAATTTTGTGAAGCAATTTCTGCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 809  
|  
QY 721 AAGGAAATATTAAGAACTACAGCAAGCAAGCTCCAGCGCACTTCTCTGATATGT 780  
DB 810 AAGGAAATATTAAGAACTACAGCAAGCAAGCTCCAGCGCACTTCTCTGATATGT 869  
|  
QY 781 ACCCCCAATATGATGACCAATGCTTAATCTGTTCAAGAGACCAAGCTTACATCTCC 840  
DB 870 ACCCCCAATATGATGACCAATGCTTAATCTGTTCAAGAGACCAAGCTTACATCTCC 929  
|  
QY 841 TTTGATGCTTTTCTGATGAGCGATGTTTAAATATGATGATGATGATGATGATGATGAT 900  
DB 930 TTTGATGCTTTTCTGATGAGCGATGTTTAAATATGATGATGATGATGATGATGATGAT 989  
|  
QY 901 GCAACACCAACATTAATTAAGCGGAAGAACAGAAAGCTTGAACAACAACTTGT 960  
DB 990 GCAACACCAACATTAATTAAGCGGAAGAACAGAAAGCTTGAACAACAACTTGT 1049  
|  
QY 961 GGAACCAAGTGTACCAAGCATTTGAGGAGCAAGAGATTTGCTGCTCTCAACGCT 1020  
DB 1050 GGAACCAAGTGTACCAAGCATTTGAGGAGCAAGAGATTTGCTGCTCTCAACGCT 1109  
|  
QY 1021 GAGCGGATTAAGACCCCAACCAAACTCTCAAGAGCGCGCAAGAGACGCTTCCCAAT 1080  
DB 1110 GAGCGGATTAAGACCCCAACCAAACTCTCAAGAGCGCGCAAGAGACGCTTCCCAAT 1169  
|  
QY 1081 AAAGTACAGAGCCCAAGCAACCCCACTTAATGCTGTAATCAAGATATACAGACGT 1140  
DB 1170 AAAGTACAGAGCCCAAGCAACCCCACTTAATGCTGTAATCAAGATATACAGACGT 1229  
|  
QY 1141 GATGAGGAGAGGAGACTGAACGCGGAGAGAAACAATGATTAAGAAAGAAAGAGAG 1200  
DB 1230 GATGAGGAGAGGAGACTGAACGCGGAGAGAAACAATGATTAAGAAAGAAAGAGAG 1289  
|  
QY 1201 AAAGTGAATCTTGAAGCTCTCTGAGCAAAATTTCTGGGTCAACCAATTAAGATG 1260  
DB 1290 AAAGTGAATCTTGAAGCTCTCTGAGCAAAATTTCTGGGTCAACCAATTAAGATG 1349  
|  
QY 1261 AAGCAATATTTGAACCTCTGAGAAATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1320  
DB 1350 AAGCAATATTTGAACCTCTGAGAAATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1409  
|  
QY 1321 AGAGTCTCATTTGCACTTATATGACAAATTTCTGTGCCATTTGTTAGTTAATTTGGAAC 1380  
DB 1410 AGAGTCTCATTTGCACTTATATGACAAATTTCTGTGCCATTTGTTAGTTAATTTGGAAC 1469  
|

QY 1381 AAAACATGTAGACAGTGTATGAGTTTGAAGTCAAGATCTAGCATATGCTCCAGCT 1440  
DB 1470 AAAACATGTAGACAGTGTATGAGTTTGAAGTCAAGATCTAGCATATGCTCCAGCT 1529  
QY 1441 CCCGCTGAGGATGTGTATCTCTCCCAAGGAAAAAGAAAGAGAAACCCGCTTGTGGCT 1500  
DB 1530 CCCGCTGAGGATGTGTATCTCTCCCAAGGAAAAAGAAAGAGAAACCCGCTTGTGGCT 1589  
QY 1501 GCACATGTCAGAAAGATACAGCTGAAAAAGACGGCTCTCTTAACATGTTTACACTAT 1560  
DB 1590 GCACATGTCAGAAAGATACAGCTGAAAAAGACGGCTCTCTTAACATGTTTACACTAT 1649  
QY 1561 CAACCTGTGTATCAATCCAGCGAGCTTGTGAAGTTGTGCTGCTTGTGTATAGCAAA 1620  
DB 1650 CAACCTGTGTATCAATCCAGCGAGCTTGTGAAGTTGTGCTGCTTGTGTATAGCAAA 1709  
QY 1621 AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGTGTCAAAAACCGCTTCCGGAGATC 1680  
DB 1710 AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGTGTCAAAAACCGCTTCCGGAGATC 1769  
QY 1681 CGCTGCAAGGACAGTGTACCAAGGAGTGCCTGCTCACTGCTGTCCGAGAGTGT 1740  
DB 1770 CGCTGCAAGGACAGTGTACCAAGGAGTGCCTGCTCACTGCTGTCCGAGAGTGT 1829  
QY 1741 GACCTGTACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATGCGGACATGTAATAATGTCTC 1800  
DB 1830 GACCTGTACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATGCGGACATGTAATAATGTCTC 1889  
QY 1801 TGCAGAAATCTGCAATATCAGCGGGCTCCAAAAAGCATATGCTGTGCGCAACATCTGAC 1860  
DB 1890 TGCAGAAATCTGCAATATCAGCGGGCTCCAAAAAGCATATGCTGTGCGCAACATCTGAC 1949  
QY 1861 GTGCGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAGAAAAATGAATTCATCTCAGAA 1920  
DB 1950 GTGCGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAGAAAAATGAATTCATCTCAGAA 2009  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAAGCTGACAGAAAGGAAAGTGTATGATAA 1980  
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAAGCTGACAGAAAGGAAAGTGTATGATAA 2069  
QY 1981 TACATGTGCACTTCTGTGTCAACTTGAACATGATTTTGTGTGTGAGAACCAACCGGCAAG 2040  
DB 2070 TACATGTGCACTTCTGTGTCAACTTGAACATGATTTTGTGTGTGAGAACCAACCGGCAAG 2129  
QY 2041 GGTAAACAATTCGTTTGTCAAAATCATTCGGTAATCCAACTGCTATGCAAAAGTTATG 2100  
DB 2130 GGTAAACAATTCGTTTGTCAAAATCATTCGGTAATCCAACTGCTATGCAAAAGTTATG 2189  
QY 2101 ATGTTTAAAGTGTATCAAGATGAGTATTTTGTCAAGAGAGCATCCAGCTGCGGAA 2160  
DB 2190 ATGTTTAAAGTGTATCAAGATGAGTATTTTGTCAAGAGAGCATCCAGCTGCGGAA 2249  
QY 2161 GAGCTGTTTTTGTATTAAGATACAGCGAGCTGATGCTGCAATATGTGCGGCAATCGAA 2220  
DB 2250 GAGCTGTTTTTGTATTAAGATACAGCGAGCTGATGCTGCAATATGTGCGGCAATCGAA 2309  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
DB 2310 AGAGAAATGAAATCCCTTGA 2330

## RESULT 3

US-10-773-302-1  
; Sequence 1, Application US/10773302  
; Publication No. US2005089880A1  
; GENERAL INFORMATION:  
; APPLICANT: Jenwein, Thomas  
; APPLICANT: Laible, Gotz  
; APPLICANT: O'Carroll, Donal  
; APPLICANT: Risenhaber, Frank  
; APPLICANT: Rea, Stephen  
; TITLE OF INVENTION: Chromatin-Regulator Genes  
; FILE REFERENCE: 0652.1670001

;; CURRENT APPLICATION NUMBER: US/10/773,302  
;; CURRENT FILING DATE: 2004-02-09  
;; PRIOR APPLICATION NUMBER: US/09/589,892  
;; PRIOR FILING DATE: 2000-06-09  
;; PRIOR APPLICATION NUMBER: US 08/945,988  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: PCT/EP96/01818  
;; PRIOR FILING DATE: 1996-05-02  
;; PRIOR APPLICATION NUMBER: DE 195 16 776.7  
;; PRIOR FILING DATE: 1995-05-10  
;; NUMBER OF SEQ ID NOS: 21  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 1  
;; LENGTH: 2600  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: (1)..(89)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (90)..(2330)  
;; FEATURE:  
;; NAME/KEY: 3'UTR  
;; LOCATION: (2331)..(2600)  
US-10-773-302-1  
Query Match 100.0%; Score 2241; DB 10; Length 2600;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGCCAGACTGGGAGAAATCTGAGAGGAGCAAGTTGTTGGCGGAAGCGTGTAAA 60  
DB 90 ATGGGCCAGACTGGGAGAAATCTGAGAGGAGCAAGTTGTTGGCGGAAGCGTGTAAA 149  
QY 61 TCAGAGTACATGCGACCTGAGACAGCTCAAGAGTTTCAAGAGCTGATGAAAGT 120  
DB 150 TCAGAGTACATGCGACCTGAGACAGCTCAAGAGTTTCAAGAGCTGATGAAAGT 209  
QY 150 TCAGAGTACATGCGACCTGAGACAGCTCAAGAGTTTCAAGAGCTGATGAAAGT 209  
QY 121 ATGTTTATGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACGAATGG 180  
DB 210 ATGTTTATGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACGAATGG 269  
QY 210 ATGTTTATGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACGAATGG 269  
DB 181 AAACAGCGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTCGCGGACT 240  
QY 270 AAACAGCGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTCGCGGACT 329  
DB 241 AGGAGGTTTGGGACCAAGTGAATTTTGCACCAAGTCAATCCATTAAGACT 300  
QY 330 AGGAGGTTTGGGACCAAGTGAATTTTGCACCAAGTCAATCCATTAAGACT 389  
DB 301 CTGATGCAATGTGCTTCAATGATCCATATATGATTTCTGCTCCCTTACAGCAATTTT 360  
QY 390 CTGATGCAATGTGCTTCAATGATCCATATATGATTTCTGCTCCCTTACAGCAATTTT 449  
DB 361 ATGTGTGAAGATGAACGTTTTCATTAACATTCCTTATATGAGGATGAAGTTTATGAT 420  
QY 450 ATGTGTGAAGATGAACGTTTTCATTAACATTCCTTATATGAGGATGAAGTTTATGAT 509  
DB 421 CAGAGATGTTCTTCAATGAAAGAACTAATAAAATATGATGAGGAAAGTACAGGGGAT 480  
QY 510 CAGAGATGTTCTTCAATGAAAGAACTAATAAAATATGATGAGGAAAGTACAGGGGAT 569  
DB 481 AGAGAAATGTGGTTTAAATGAAGAAATTTTGTGAGTGTGAAATGCCCTTGTGCA 540  
QY 570 AGAGAAATGTGGTTTAAATGAAGAAATTTTGTGAGTGTGAAATGCCCTTGTGCA 629  
DB 541 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 630 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
DB 601 AAAGATCTGAGGATCCCGAGATGATTAAGAAAGCGGCACTCGGAAATTTCTCTCT 660

Db 690 AAGATCTGAGGATCAACGAGATGATAAAGAAAGCCGACCTCGAAATTCCTCT 749  
 Qy 661 GATTAATTTTGAAGCCATTTCTCAATGTTTCCGATTAAGGCAACGCAAAACCTA 720  
 Db 750 GATTAATTTTGAAGCCATTTCTCAATGTTTCCGATTAAGGCAACGCAAAACCTA 809  
 Qy 721 AAGGAAAAATTAAGAACTCAACGCAACGAGCTCCAGGCGCACTTCTCTGAAATG 780  
 Db 810 AAGGAAAAATTAAGAACTCAACGCAACGAGCTCCAGGCGCACTTCTCTGAAATG 869  
 Qy 781 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTAGAGAGAGCAAACTTACCTCC 840  
 Db 870 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTAGAGAGAGCAAACTTACCTCC 929  
 Qy 841 TTTCAATGCTTTTCTGTAGGCGATGTTTAAATATGATGCTTCTTCAATCTTTTCA 900  
 Db 930 TTTCAATGCTTTTCTGTAGGCGATGTTTAAATATGATGCTTCTTCAATCTTTTCA 989  
 Qy 901 GCAACACCCACACTATTAAGGCGAAAGCAACAGAAAGCTCTAGACAAACCTTGT 960  
 Db 990 GCAACACCCACACTATTAAGGCGAAAGCAACAGAAAGCTCTAGACAAACCTTGT 1049  
 Qy 961 GAGACCAAGTGTATCAGCAATTTGAGGAGCAAAAGATTTGCTGCTCTCAACGCT 1020  
 Db 1050 GAGACCAAGTGTATCAGCAATTTGAGGAGCAAAAGATTTGCTGCTCTCAACGCT 1109  
 Qy 1021 GAGCGATTAAGACCCCAACCAACGTCAGAGGCGCGAGAGAGAGAGCTTCCCAT 1080  
 Db 1110 GAGCGATTAAGACCCCAACCAACGTCAGAGGCGCGAGAGAGAGAGCTTCCCAT 1169  
 Qy 1081 AACAGTAGAGGCGCGACCCCAACCAATTAATGCTGGAATCAAGAGTACAGACGT 1140  
 Db 1170 AACAGTAGAGGCGCGACCCCAACCAATTAATGCTGGAATCAAGAGTACAGACGT 1229  
 Qy 1141 GATAGGGAAGCAAGGAGCTGAAAAGGAGGAGAGAAATATTAAGAGAGAGAGAG 1200  
 Db 1230 GATAGGGAAGCAAGGAGCTGAAAAGGAGGAGAGAAATATTAAGAGAGAGAGAG 1289  
 Qy 1201 AAGATGAAATCTTCAAGCTCTCTGAAAGCAATTTCTCGGTCTCAACCAATTAAGATG 1260  
 Db 1290 AAGATGAAATCTTCAAGCTCTCTGAAAGCAATTTCTCGGTCTCAACCAATTAAGATG 1349  
 Qy 1261 AAGCCAAATATTAAGCTCTCTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1320  
 Db 1350 AAGCCAAATATTAAGCTCTCTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1409  
 Qy 1321 AAGATCTCTATGAGCACTTACTATGCAATTTCTGAGCAATTTGAGTAAATGGGACC 1380  
 Db 1410 AAGATCTCTATGAGCACTTACTATGCAATTTCTGAGCAATTTGAGTAAATGGGACC 1469  
 Qy 1381 AAAACATGTAGAAGGTGTATGATTTAGAGTCAAAAGATCTAGCATCATACCTCAGCT 1440  
 Db 1470 AAAACATGTAGAAGGTGTATGATTTAGAGTCAAAAGATCTAGCATCATACCTCAGCT 1529  
 Qy 1441 CCGCGTAGAGATGTGATCTCTCTCAAGAGAAAGAAAGAGAAACACCGGTGTGGGCT 1500  
 Db 1530 CCGCGTAGAGATGTGATCTCTCTCAAGAGAAAGAAAGAGAAACACCGGTGTGGGCT 1589  
 Qy 1501 GCAACTGTGATCATTCACGAGGAGCTTGTGACAGTGTGCGCTTGTGTATGACAA 1560  
 Db 1590 GCAACTGTGATCATTCACGAGGAGCTTGTGACAGTGTGCGCTTGTGTATGACAA 1649  
 Qy 1561 CAACCTGTGATCATTCACGAGGAGCTTGTGACAGTGTGCGCTTGTGTATGACAA 1620  
 Db 1650 CAACCTGTGATCATTCACGAGGAGCTTGTGACAGTGTGCGCTTGTGTATGACAA 1709  
 Qy 1621 AATTTTGTGAAAGTTTGTCAATGTACTGTAGAGTGTCAAAACCGCTTCCGGATGC 1680  
 Db 1710 AATTTTGTGAAAGTTTGTCAATGTACTGTAGAGTGTCAAAACCGCTTCCGGATGC 1769  
 Qy 1681 CGCTGCAAGCAAGTGAACACCAAGAGTGCCGCTGCTACTGCTGTCCGAGAGTGT 1740  
 Db 1770 CGCTGCAAGCAAGTGAACACCAAGAGTGCCGCTGCTACTGCTGTCCGAGAGTGT 1829

Qy 1741 GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATGGGACAGTAAATGTGTCC 1800  
 Db 1830 GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATGGGACAGTAAATGTGTCC 1889  
 Qy 1801 TGCAGAACTGCAAGTATTCAGCGGGGCTCAAAAAGATCTATTTCTGGCACATCTGAC 1860  
 Db 1890 TGCAGAACTGCAAGTATTCAGCGGGGCTCAAAAAGATCTATTTCTGGCACATCTGAC 1949  
 Qy 1861 GTGGCAGGCTGGGGATTTTATTCAAAATCTCTGTGCAAGAAAATGAATTCATCTCAGA 1920  
 Db 1950 GTGGCAGGCTGGGGATTTTATTCAAAATCTCTGTGCAAGAAAATGAATTCATCTCAGA 2009  
 Qy 1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAAAGTATGATATA 1980  
 Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAAAGTATGATATA 2069  
 Qy 1981 TACATGTGAGCTTCTGTTCACCTTGAAACATGATTTTGTGTGATGCAACCCGCAAG 2040  
 Db 2070 TACATGTGAGCTTCTGTTCACCTTGAAACATGATTTTGTGTGATGCAACCCGCAAG 2129  
 Qy 2041 GGTAAACAAATTCGTTTTCGAATCATTCGGTAAATCCAACTGCTATGCAAAAGTATG 2100  
 Db 2130 GGTAAACAAATTCGTTTTCGAATCATTCGGTAAATCCAACTGCTATGCAAAAGTATG 2189  
 Qy 2101 ATGTTTAAGGTATCAGAGTAGATATTTTGTGCAAGAGCCATCCAGCTGAGCGCA 2160  
 Db 2190 ATGTTTAAGGTATCAGAGTAGATATTTTGTGCAAGAGCCATCCAGCTGAGCGCA 2249  
 Qy 2161 GAGCTGTTTGTATTAAGATACAGCAGGCTGATGCTGTAAGTATGTGGCATGCA 2220  
 Db 2250 GAGCTGTTTGTATTAAGATACAGCAGGCTGATGCTGTAAGTATGTGGCATGCA 2309  
 Qy 2221 AGAGAAATGGAATCCCTTGA 2241  
 Db 2310 AGAGAAATGGAATCCCTTGA 2330

RESULT 4  
 US-10-723-860-5458  
 ; Sequence 5458, Application US/10723860  
 ; Publication No. US20040253606A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Aziz, Metabpa  
 ; APPLICANT: Ginsburg, Wendy M.  
 ; APPLICANT: Zlotnik, Albert  
 ; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
 ; FILE REFERENCE: 05882, 0193, NPUS01  
 ; CURRENT APPLICATION NUMBER: US/10/723, 860  
 ; PRIOR FILING DATE: 2003-11-26  
 ; PRIOR APPLICATION NUMBER: 60/429, 739  
 ; NUMBER OF SEQ ID NOS: 8393  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO: 5458  
 ; LENGTH: 2711  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-723-860-5458

Query Match 100.0%; Score 2241; DB 9; Length 2711;  
 Best Local Similarity 100.0%; P-Val. No. 0;  
 Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ATGGGCGAGACTGGGAGAAATCTGAGAGGAGCAAGTTTGTGCGGAGAGCTGTAAAG 60  
 Db 167 ATGGGCGAGACTGGGAGAAATCTGAGAGGAGCAAGTTTGTGCGGAGAGCTGTAAAG 226  
 Qy 61 TCAAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAGTAAAGAT 120  
 Db 227 TCAAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCAGCTGATGAGTAAAGAT 286



Qy	121	ATGTTTGGTTCCTCCATTCGACACAAAATTTTGGAAAGAAACGAAATCTTAAACCAAGATNG	180
Db	287	ATGTTTGGTTCCTCCATTCGACACAAAATTTTGGAAAGAAACGAAATCTTAAACCAAGATNG	346
Qy	181	AAAACAGCAAGATATACAGCTGTGTGCACATCTGTGACTTCTGTGAGCTCATTTGCGGGAGCT	240
Db	347	AAAACAGCAAGATATACAGCTGTGTGCACATCTGTGACTTCTGTGAGCTCATTTGCGGGAGCT	406
Qy	241	AGGAGTGTTCGGTGCACCAAGTACTGTGATTTTTCACACACAGTCAATCCATTAAAGACT	300
Db	407	AGGAGTGTTCGGTGCACCAAGTACTGTGATTTTTCACACACAGTCAATCCATTAAAGACT	466
Qy	301	CTGAATGAGTGTGCTTCAGTACCAATAATATCTTGATCTGCTGCCCTACAGCAAAATTTT	360
Db	467	CTGAATGAGTGTGCTTCAGTACCAATAATATCTTGATCTGCTGCCCTACAGCAAAATTTT	526
Qy	361	ATGTGTGAAGATGAAACGTGTTTACATPAACATTCCTTATATGAGAGATGAAGTTTATGAT	420
Db	527	ATGTGTGAAGATGAAACGTGTTTACATPAACATTCCTTATATGAGAGATGAAGTTTATGAT	586
Qy	421	CAGATGTGACTTTCATTGAAGAACTAATTAATAAAATTAATGATGAGAAAGTACACGGGAT	480
Db	587	CAGATGTGACTTTCATTGAAGAACTAATTAATAAAATTAATGATGAGAAAGTACACGGGAT	646
Qy	481	AGAGAAATGTGGGTTTATTAATGATGAAATTTTGTGTGAGTGTGTGATATGCCCTTGTCAA	540
Db	647	AGAGAAATGTGGGTTTATTAATGATGAAATTTTGTGTGAGTGTGTGATATGCCCTTGTCAA	706
Qy	541	TATATATGATGATGACGATGATGATGATGACGATCTGTGAAGAAAGAGAAAGAACAG	600
Db	707	TATATATGATGATGACGATGATGATGATGACGATCTGTGAAGAAAGAGAAAGAACAG	766
Qy	601	AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCGGCCACCTCGAAATTTTCTTCT	660
Db	767	AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCGGCCACCTCGAAATTTTCTTCT	826
Qy	661	GATATAAATTTTGTGAAGCCATTTTCTCTCATATGTTTCCAGATAAGGGCACAGCAGAAAGATA	720
Db	827	GATATAAATTTTGTGAAGCCATTTTCTCTCATATGTTTCCAGATAAGGGCACAGCAGAAAGATA	886
Qy	721	AAGGAAAAATTAATAAGAACTACCGGAAACAGGAGCTCCACGGCGCATTCCTCTGAATGT	780
Db	887	AAGGAAAAATTAATAAGAACTACCGGAAACAGGAGCTCCACGGCGCATTCCTCTGAATGT	946
Qy	781	ACCCCCACATAGATGACCAATATGCTAAATCTGTTCAGAGAGACAAAGCTTACATCTCC	840
Db	947	ACCCCCACATAGATGACCAATATGCTAAATCTGTTCAGAGAGACAAAGCTTACATCTCC	1006
Qy	841	TTTTCATAGGCTTTTCTGTAGGCGATGTTTAAATATGATCTGCTTCTTACATCTCTTTTCA	900
Db	1007	TTTTCATAGGCTTTTCTGTAGGCGATGTTTAAATATGATCTGCTTCTTACATCTCTTTTCA	1066
Qy	901	GCAACACCCCAACATTAATAGCGGAGAAACACAGAAACAGCTTACAGCAACAACTTGT	960
Db	1067	GCAACACCCCAACATTAATAGCGGAGAAACACAGAAACAGCTTACAGCAACAACTTGT	1126
Qy	961	GGACCAACGTGTACCAAGATTTTGGAGGAGCAAAAGATTTGCTGTGCTCTCACCGCT	1020
Db	1127	GGACCAACGTGTACCAAGATTTTGGAGGAGCAAAAGATTTGCTGTGCTCTCACCGCT	1186
Qy	1021	GAGCGGATTAAGAATCCCAACAAACGTCCAGAGAGCGCAGAAAGAGACGGCTTCCAT	1080
Db	1187	GAGCGGATTAAGAATCCCAACAAACGTCCAGAGAGCGCAGAAAGAGACGGCTTCCAT	1246
Qy	1081	AAACGTACAGGCCCAAGACCCCAACATTAATGTGCTGGAATCAAAAGATATACAGACGT	1140
Db	1247	AAACGTACAGGCCCAAGACCCCAACATTAATGTGCTGGAATCAAAAGATATACAGACGT	1306
Qy	1141	GATAGGGAAGACAGGACTGAAGACGGGGGAGAGAACATGATTAAGAAGAAAGAGAGAG	1200
Db	1307	GATAGGGAAGACAGGACTGAAGACGGGGGAGAGAACATGATTAAGAAGAAAGAGAGAG	1366
Qy	1201	AAAGTGAACCTTGAGCTCCTCTGAGACAAATTCCTGGGTCTCAACACCAATTAAGATG	1260

Db	1367	AAAGATGAACTTCTGAGCTCCTCTGAAGCAAAATTCCTGGTGTCAACACCAATPAAAGATG	1426
Qy	1261	AAGCCAAATATGAACTCTTGAGATGTGAGTGTGAGTGTGTGAAGCTTCAATGTTT	1320
Db	1427	AAGCCAATATTGGAACCTCTCGAAGATGTGAGTGTGTGTGTGAAGCTCAATGTTT	1486
Qy	1321	AGAGTCTCATTTGGCACTTATCTATGACAAATTTTCGTGGCATTTGCTAGTTAATTTGGACC	1380
Db	1487	AGAGTCTCATTTGGCACTTATCTATGACAAATTTTCGTGGCATTTGCTAGTTAATTTGGACC	1546
Qy	1381	AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAAATCTAGCATATAGCTCCAGCT	1440
Db	1547	AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAAATCTAGCATATAGCTCCAGCT	1606
Qy	1441	CCCGCTGAGAGATGTGATACTCTCTCCAGGAAAAAGAAAGGAAACACCGGTGTGGGCT	1500
Db	1607	CCCGCTGAGAGATGTGATACTCTCTCCAGGAAAAAGAAAGGAAACACCGGTGTGGGCT	1666
Qy	1501	GCAACATGCGAAAGATPACAGCTGAAAAAAGGAGCGGCTCCCTAACCAATGTTTCAACTAT	1560
Db	1667	GCAACATGCGAAAGATPACAGCTGAAAAAAGGAGCGGCTCCCTAACCAATGTTTCAACTAT	1726
Qy	1561	CAACCCGTGATCAATCCAGCGGACGCTTGTGACAGTTTGTGACCCCTGTGTGATAGCAAA	1620
Db	1727	CAACCCGTGATCAATCCAGCGGACGCTTGTGACAGTTTGTGACCCCTGTGTGATAGCAAA	1786
Qy	1621	AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAGAGTGTCAAAAACCGCTTTCCGGGATGC	1680
Db	1787	AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAGAGTGTCAAAAACCGCTTTCCGGGATGC	1846
Qy	1681	CGCTGCGAAAGACAGTGTGCAACCAACAGTGGCCGGTACCTACCGGCTGTCCGAGAGTGT	1740
Db	1847	CGCTGCGAAAGACAGTGTGCAACCAACAGTGGCCGGTACCTACCTACCAATGTATTCAGAGTGT	1906
Qy	1741	GACCCCTGACCTGTCTTACTTGTGTGAGCCGCTGACCAATTTGGGACAGTAAAAATGTGTCC	1800
Db	1907	GACCCCTGACCTGTCTTACTTGTGTGAGCCGCTGACCAATTTGGGACAGTAAAAATGTGTCC	1966
Qy	1801	TGCAGAAATGTGCAATATTCAGCCGGGCTCCAAAAAGCATATTTGCTGTGCAACATCTGAC	1860
Db	1967	TGCAGAAATGTGCAATATTCAGCCGGGCTCCAAAAAGCATATTTGCTGTGCAACATCTGAC	2026
Qy	1861	GTGGCAGAGCTGGGGGATTTTATTAAGAAATCCTGTGCAAAAAATGAATTCATCTCAGAA	1920
Db	2027	GTGGCAGAGCTGGGGGATTTTATTAAGAAATCCTGTGCAAAAAATGAATTCATCTCAGAA	2086
Qy	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTGTATGATAA	1980
Db	2087	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTGTATGATAA	2146
Qy	1981	TACATGTGCAAGCTTCTGTTCACACTTGAACATATTTTGTGTGATGCAACCCGCAAG	2040
Db	2147	TACATGTGCAAGCTTCTGTTCACACTTGAACATATTTTGTGTGATGCAACCCGCAAG	2206
Qy	2041	GGTAAACAAATGTGTTTGTGCAAAATCATTCGGTAAATCAAACTGCTATGCAAAAAGTATG	2100
Db	2207	GGTAAACAAATGTGTTTGTGCAAAATCATTCGGTAAATCAAACTGCTATGCAAAAAGTATG	2266
Qy	2101	ATGTGTTAACGGTGTATCAAGAGATAGATATTTTGTCCAAAGAGGCAATCCAGACTGTGGAA	2160
Db	2267	ATGTGTTAACGGTGTATCAAGAGATAGATATTTTGTCCAAAGAGGCAATCCAGACTGTGGAA	2326
Qy	2161	GAGCTGTTTTTGTATTCAGATACAGCCAGGCTGATGCCCTGAAGTATGTCCGCAATGAA	2220
Db	2327	GAGCTGTTTTTGTATTCAGATACAGCCAGGCTGATGCCCTGAAGTATGTCCGCAATGAA	2386
Qy	2221	AGAGAAATGAAAAATCCCTTGA 2241	
Db	2387	AGAGAAATGAAAAATCCCTTGA 2407	

US-10-153-668-231  
; Sequence 231, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAMA, Kenya  
; TITLE OF INVENTION: STATE Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10153,668  
; PRIORITY FILING DATE: 2002-05-24  
; PRIORITY APPLICATION NUMBER: US 60/293,172  
; PRIORITY FILING DATE: 2001-05-25  
; PRIORITY APPLICATION NUMBER: US 60/316,031  
; PRIORITY FILING DATE: 2001-08-31  
; PRIORITY APPLICATION NUMBER: US 60/328,403  
; PRIORITY FILING DATE: 2001-10-12  
; PRIORITY APPLICATION NUMBER: JP 2001-157043  
; PRIORITY FILING DATE: 2001-05-25  
; PRIORITY APPLICATION NUMBER: JP 2001-260681  
; PRIORITY FILING DATE: 2001-08-30  
; PRIORITY APPLICATION NUMBER: JP 2001-313175  
; PRIORITY FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 231  
; LENGTH: 2558  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (58) .. (2295)  
US-10-153-668-231

Query Match 99.8%; Score 2236.2; DB 6; Length 2558;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGCGACAGTGGGAGAAATCTGAGAAAGGACAGTTGTTGGCGGAGCGGTGTAAGA 60  
DB 58 ATGGGCGACAGTGGGAGAAATCTGAGAAAGGACAGTTGTTGGCGGAGCGGTGTAAGA 117  
QY 61 TCGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCTGATGTAAGTAAGAT 120  
DB 118 TCGAGTACATGCGACTGAGACAGCTCAAGAGTTTCAAGCTGATGTAAGTAAGAT 177  
QY 121 ATGTTAGTTCCAAATGTCAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATG 180  
DB 178 ATGTTAGTTCCAAATGTCAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGATG 237  
QY 181 AAACAGCGAAGATACAGCTGTGCAATCTGATCTTGTGAGCTCATTTGGCGGAGCT 240  
DB 238 AAACAGCGAAGATACAGCTGTGCAATCTGATCTTGTGAGCTCATTTGGCGGAGCT 297  
QY 241 AGGAGTGTTCGTTGACAGTGAATTTTCCAAACCAAGTCAATCCATTAAAGCT 300  
DB 298 AGGAGTGTTCGTTGACAGTGAATTTTCCAAACCAAGTCAATCCATTAAAGCT 357  
QY 301 CTGATGCACTGCTTCAATACCCCAATTAATTTCTGCTCCCTTACAGAGAAATTT 360  
DB 358 CTGATGCACTGCTTCAATACCCCAATTAATTTCTGCTCCCTTACAGAGAAATTT 417  
QY 361 ATGTGGAAGATGAATCTTTTACATTAATCTTATATGAGAGTGAAGTTTAT 420  
DB 418 ATGTGGAAGATGAATCTTTTACATTAATCTTATATGAGAGTGAAGTTTAT 477  
QY 421 CAGGATGTAATTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 480  
DB 478 CAGGATGTAATTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 537  
QY 481 AAGAGATGAGGTTTAAATGATGAATTTTGTGAGATGAGTGAATGCTGTGCTCA 540

DB 538 AAGAGATGAGGTTTAAATGATGAATTTTGTGAGATGAGTGAATGCTGTGCTCA 597  
QY 541 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 598 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 601 AAAGATCTGAGAGTCAACGAGATGTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660  
DB 658 AAAGATCTGAGAGTCAACGAGATGTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 717  
QY 661 GATTAATTTTGAAGCAATTTCTCAATGTTTCAAGATTAAGGCAACGAGAGACTA 720  
DB 718 GATTAATTTTGAAGCAATTTCTCAATGTTTCAAGATTAAGGCAACGAGAGACTA 777  
QY 721 AAGGAAAAATTAAGAACTCAACGAGCTCCAGGCGCACTTCTCTTAAT 780  
DB 778 AAGGAAAAATTAAGAACTCAACGAGCTCCAGGCGCACTTCTCTTAAT 837  
QY 781 ACCCCCAATGATGAGCAAAATGCTAAATGTTTCAAGAGAGCAAAAGCTTACACTCC 840  
DB 838 ACCCCCAATGATGAGCAAAATGCTAAATGTTTCAAGAGAGCAAAAGCTTACACTCC 897  
QY 841 TTTCAATGCTTTTCTGATGAGGATGTTTAAATGATGATGATGATGATGATGATGAT 900  
DB 898 TTTCAATGCTTTTCTGATGAGGATGTTTAAATGATGATGATGATGATGATGATGAT 957  
QY 901 GCAACACCAACATTAATTAAGGAGAAACAGAAACAGCTCTTGAACAACAACTTGT 960  
DB 958 GCAACACCAACATTAATTAAGGAGAAACAGAAACAGCTCTTGAACAACAACTTGT 1017  
QY 961 GGACCAAGTGTACACAGATTTGAGAGGAGCAAGATTTGCTGCTCTCACCGCT 1020  
DB 1018 GGACCAAGTGTACACAGATTTGAGAGGAGCAAGATTTGCTGCTCTCACCGCT 1077  
QY 1021 GAGCGAATTAAGACCCCAACCAAAAGTCCAGAGGCGCGAGAGAGAGAGCGGTTCCAT 1080  
DB 1078 GAGCGAATTAAGACCCCAACCAAAAGTCCAGAGGCGCGAGAGAGAGAGCGGTTCCAT 1137  
QY 1081 AACGATGACAGGCGCCAGACCCCACTTAATGCTGAGATCAAAAGATCAACAGCT 1140  
DB 1138 AACGATGACAGGCGCCAGACCCCACTTAATGCTGAGATCAAAAGATCAACAGCT 1197  
QY 1141 GATGAGGAGAGAGGAGCTGAACCGGGGAGAGAACATGATTAAGAAAGAGAGAG 1200  
DB 1198 GATGAGGAGAGAGGAGCTGAACCGGGGAGAGAACATGATTAAGAAAGAGAGAG 1257  
QY 1201 AAAGATGAATCTTCAAGCTCTCTGAGCAAAATTTCTGCTGCAAAACCAATTAAGAT 1260  
DB 1258 AAAGATGAATCTTCAAGCTCTCTGAGCAAAATTTCTGCTGCAAAACCAATTAAGAT 1317  
QY 1261 AAGCAATTAATTAAGCTCTCTGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1320  
DB 1318 AAGCAATTAATTAAGCTCTCTGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1377  
QY 1321 AAGTCTCTCAATTTGAGCTTAATGAACAATTTCTGCTGCAATTTGAGAGCC 1380  
DB 1378 AAGTCTCTCAATTTGAGCTTAATGAACAATTTCTGCTGCAATTTGAGAGCC 1437  
QY 1381 AAAACATGATGACAGGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1438 AAAACATGATGACAGGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1497  
QY 1441 CCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
DB 1498 CCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557  
QY 1501 GCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
DB 1558 GCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1617  
QY 1561 CAACCTGTGATCAATCCAGGAGCTTGTGACAGTTGTGCTCTTGTGTGATGACAA 1620  
DB 1618 CAACCTGTGATCAATCCAGGAGCTTGTGACAGTTGTGCTCTTGTGTGATGACAA 1677

OY 1621 AATTTTGTGAAAAGTTTGTCAATGTAGTTCAGAGTGTCAAAAACCGCTTCCGGAGATGC 1680  
DB 1678 AATTTTGTGAAAAGTTTGTCAATGTAGTTCAGAGTGTCAAAAACCGCTTCCGGAGATGC 1737  
OY 1681 CGCTGCAAGCACTGTGCAACCAAGAGTGCCTGTCTACCTGTCTGTCCGAGAGTGT 1740  
DB 1738 CGCTGCAAGCACTGTGCAACCAAGAGTGCCTGTCTACCTGTCTGTCCGAGAGTGT 1797  
OY 1741 GACCTTGACCTCTGTCTACTTTGTGAGACCGCTGACCAATTGGGACAGTAAATGTC 1800  
DB 1798 GACCTTGACCTCTGTCTACTTTGTGAGACCGCTGACCAATTGGGACAGTAAATGTC 1857  
OY 1801 TGCAGAACTGTGCAATTCAGCGGGCTCAAAAAGCATCTATTGTCTGTGCAACCTGTAC 1860  
DB 1858 TGCAGAACTGTGCAATTCAGCGGGCTCAAAAAGCATCTATTGTCTGTGCAACCTGTAC 1917  
OY 1861 GTGGCAGGCTGGGGATTTTATCAAGATCTGTGCAAAAATGAAATTCATCTCAGAA 1920  
DB 1918 GTGGCAGGCTGGGGATTTTATCAAGATCTGTGCAAAAATGAAATTCATCTCAGAA 1977  
OY 1921 TACTGTGAGAGATTTATTTTCCAAAGTGAAGCTGACAGAAAGGGAAGTGTATGATAA 1980  
DB 1978 TACTGTGAGAGATTTATTTTCCAAAGTGAAGCTGACAGAAAGGGAAGTGTATGATAA 2037  
OY 1981 TACATGTGACAGCTTCTGTTCAACTTGAAACAATGATTTTGTGTGATGCAACCCGCAAG 2040  
DB 2038 TACATGTGACAGCTTCTGTTCAACTTGAAACAATGATTTTGTGTGATGCAACCCGCAAG 2097  
OY 2041 GGTAAACAATTCGTTTGGCAATCAATCGGTAAATCCAACTGCTATGCAAAAGTTATG 2100  
DB 2098 GGTAAACAATTCGTTTGGCAATCAATCGGTAAATCCAACTGCTATGCAAAAGTTATG 2157  
OY 2101 ATGTGTAACGGTATCAAGATAGATATTTTGTGCAAGAGCCATCGAATGCGCA 2160  
DB 2158 ATGTGTAACGGTATCAAGATAGATATTTTGTGCAAGAGCCATCGAATGCGCA 2217  
OY 2161 GAGCTGTTTTTGTATTAAGATACAGCCAGGCTGATGCCCTGAAGTATGTGCGCATGAA 2220  
DB 2218 GAGCTGTTTTTGTATTAAGATACAGCCAGGCTGATGCCCTGAAGTATGTGCGCATGAA 2277  
OY 2221 AGAGAAATGGAATCCCTTGA 2241  
DB 2278 AGAGAAATGGAATCCCTTGA 2298

RESULT 6  
US-10-172-118-1043  
Sequence 1043, Application US/10172118  
Publication No. US20030224374A1  
GENERAL INFORMATION:  
APPLICANT: Dai, Hongyue  
APPLICANT: He, Yudong  
APPLICANT: Linsley, Peter  
APPLICANT: Mao, Mao  
APPLICANT: Roberts, Chris  
APPLICANT: Van 't Veer, Laura  
APPLICANT: Van de Vijver, Marc  
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
FILE REFERENCE: 9301-175-999  
CURRENT APPLICATION NUMBER: US/10/172,118  
CURRENT FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: 60/380,770  
PRIOR FILING DATE: 2002-05-14  
NUMBER OF SEQ ID NOS: 2699  
SEQ ID NO 1043  
LENGTH: 2576  
TYPE: DNA  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
DATABASE ACCESSION NUMBER: NM\_004456  
DATABASE ENTRY DATE: 2001-06-18

US-10-172-118-1043

Query Match 99.8%; Score 2236.2; DB 7; Length 2576;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2236; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGGGCGAGACTGGGAAGAAATCTGAGAGGACCAAGTTTGTGGCGGAGCGGTATAAA 60  
DB 58 ATGGGCGAGACTGGGAAGAAATCTGAGAGGACCAAGTTTGTGGCGGAGCGGTATAAA 117  
OY 61 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTGAGAGAGCTGATGAAGTAAAGT 120  
DB 118 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTGAGAGAGCTGATGAAGTAAAGT 177  
OY 121 ATGTTAGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGATGG 180  
DB 178 ATGTTAGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGATGG 237  
OY 181 AAACAGGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGCGGAGACT 240  
DB 238 AAACAGGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGCGGAGACT 297  
OY 241 AGGAGGTTCGGTGAACAGTGAAGTGAATTTTCCACACAGATCCCATTAAGACT 300  
DB 298 AGGAGGTTCGGTGAACAGTGAAGTGAATTTTCCACACAGATCCCATTAAGACT 357  
OY 301 CTGATGCAAGTTGCTTCAATACCATTAATGATTTCTTGTCTCCCTTACAGAGATTTT 360  
DB 358 CTGATGCAAGTTGCTTCAATACCATTAATGATTTCTTGTCTCCCTTACAGAGATTTT 417  
OY 361 ATGTGTGAAGATGAAATCTGTTTATCATTAATCTTATATGGAGATGAAGTTTATGAT 420  
DB 418 ATGTGTGAAGATGAAATCTGTTTATCATTAATCTTATATGGAGATGAAGTTTATGAT 477  
OY 421 CAGATGCTACTTTCATTGAAGAACTAATAAATTTATGATGAGAAAGTACACGGGAT 480  
DB 478 CAGATGCTACTTTCATTGAAGAACTAATAAATTTATGATGAGAAAGTACACGGGAT 537  
OY 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGAATGCCCTGTGCA 540  
DB 538 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGAATGCCCTGTGCA 597  
OY 541 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
DB 598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
OY 601 AAAGATCTGAGAGTACCCGAGATGATTAAGAAAGCCGCCCATCTCGGAATTTCTTCT 660  
DB 658 AAAGATCTGAGAGTACCCGAGATGATTAAGAAAGCCGCCCATCTCGGAATTTCTTCT 717  
OY 661 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGACAGCAAGAAAGACTA 720  
DB 718 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGACAGCAAGAAAGACTA 777  
OY 721 AAAGAAAAATTAAGAAATCTCAAGAAACAGAGCTCCAGGCGGACTTCTCTGAATGT 780  
DB 778 AAAGAAAAATTAAGAAATCTCAAGAAACAGAGCTCCAGGCGGACTTCTCTGAATGT 837  
OY 781 ACCCCACATAGATGACCAATGCTAAATCTGTTGAGAGAGCAAGGTTTACATCTCC 840  
DB 838 ACCCCACATAGATGACCAATGCTAAATCTGTTGAGAGAGCAAGGTTTACATCTCC 897  
OY 841 TTTCAATGCTTTTCTGTAGCGATGTTTAAATATGATGCTGTTCTTCAATCTTTTCAAT 900  
DB 898 TTTCAATGCTTTTCTGTAGCGATGTTTAAATATGATGCTGTTCTTCAATCTTTTCAAT 957  
OY 901 GCAACACCAACATTAATAAGCGAAGAACACAGAAACAGCTTATGACAAACCTTGT 960  
DB 958 GCAACACCAACATTAATAAGCGAAGAACACAGAAACAGCTTATGACAAACCTTGT 1017  
OY 961 GGAACACAGTGTATCAGCATTTTGAAGGAGCAAAAGATTTGTGCTGTCTCACCGCT 1020  
DB 1018 GGAACACAGTGTATCAGCATTTTGAAGGAGCAAAAGATTTGTGCTGTCTCACCGCT 1077



[illegible]

QY	1201	AAAGATGAACTTCGAGCTCCTCGAAGCAAAATTCGAGTCAAAACCAATPAAGATG	1260
Db	1258	AAAGATGAACTTCGAGCTCCTCGAAGCAAAATTCGAGTCAAAACCAATPAAGATG	1317
QY	1261	AAGCAAAATATGAACTCCTGAGAAATGAGAGTGAAGTGAAGCTGAAGCTCAATGTTT	1320
Db	1318	AAGCAAAATATGAACTCCTGAGAAATGAGAGTGAAGTGAAGCTGAAGCTCAATGTTT	1377
QY	1321	AGAGTCTCATTTGGCACTTACTATGACAAATTTCTGTGACCATTGCTAGTTAAATTTGGACC	1380
Db	1378	AGAGTCTCATTTGGCACTTACTATGACAAATTTCTGTGACCATTGCTAGTTAAATTTGGACC	1437
QY	1381	AAAACATGTAGACAGAGTATGAGTTAAGTCAAGAAATCTAGACATATAGCTCCAGCT	1440
Db	1438	AAAACATGTAGACAGAGTATGAGTTAAGTCAAGAAATCTAGACATATAGCTCCAGCT	1497
QY	1441	CCGCTGAGAGATGTGATTAATCCTCCAGAGAAAAAGAGGAAACACCGGTGTGGGCT	1500
Db	1498	CCGCTGAGAGATGTGATTAATCCTCCAGAGAAAAAGAGGAAACACCGGTGTGGGCT	1557
QY	1501	GCACATGCGAGAAAGATACAGCTGAAAAAAGAGGGCTCCCTTAACCATGTTTACAATAT	1560
Db	1558	GCACATGCGAGAAAGATACAGCTGAAAAAAGAGGGCTCCCTTAACCATGTTTACAATAT	1617
QY	1561	CAACCTGTGATCATCCACGCGAGCCTTGTGACAGTTTCGTGCCCTGTGTATGACACA	1620
Db	1618	CAACCTGTGATCATCCACGCGAGCCTTGTGACAGTTTCGTGCCCTGTGTATGACACA	1677
QY	1621	AATTTTGTGAAAAAGTTTGTGCATATGTATCAGAGTCAAAACCGCTTCCGGGATGC	1680
Db	1678	AATTTTGTGAAAAAGTTTGTGCATATGTATCAGAGTCAAAACCGCTTCCGGGATGC	1737
QY	1681	CGCTGCAAAAGACAGTGCACAACAAGCAGTGCACCGGTCTACCTGCTGTCGAGAGTGT	1740
Db	1738	CGCTGCAAAAGACAGTGCACAACAAGCAGTGCACCGGTCTACCTGCTGTCGAGAGTGT	1797
QY	1741	GACCTTGACCTCTGTCTTAACTTGTGAGCCGCTGACCATTTGGGACAGTAAAAATGTGTCC	1800
Db	1798	GACCTTGACCTCTGTCTTAACTTGTGAGCCGCTGACCATTTGGGACAGTAAAAATGTGTCC	1857
QY	1801	TGCAAGAAATGCGAGTATTCAGCGGGGGCTCCAAAAGACATCTATGCTGGGACCATCTGCAC	1860
Db	1858	TGCAAGAAATGCGAGTATTCAGCGGGGGCTCCAAAAGACATCTATGCTGGGACCATCTGCAC	1917
QY	1861	GTGGCAGAGCTGGGGGATTTTATTCAAAGATCTGTGCAAAAAATGAATTCATCTAGAA	1920
Db	1918	GTGGCAGAGCTGGGGGATTTTATTCAAAGATCTGTGCAAAAAATGAATTCATCTAGAA	1977
QY	1921	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAGTATGATPAA	1980
Db	1978	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAGTATGATPAA	2037
QY	1981	TACATGTGAGCTTCTGTTCAACTTGAAACAATGATTTGTGTGTGATGACACCCGCAAG	2040
Db	2038	TACATGTGAGCTTCTGTTCAACTTGAAACAATGATTTGTGTGTGATGACACCCGCAAG	2097
QY	2041	GGTAAACAATTCGTTTGGCAATCATTCGGTAAATCAAACTGCTATGCAAAAATTTATG	2100
Db	2098	GGTAAACAATTCGTTTGGCAATCATTCGGTAAATCAAACTGCTATGCAAAAATTTATG	2157
QY	2101	ATGGTTAAACGGTATACAGGATATGATTTTGGCAAGAGACCATCCGATCGGGGAA	2160
Db	2158	ATGGTTAAACGGTATACAGGATATGATTTTGGCAAGAGACCATCCGATCGGGGAA	2217
QY	2161	GAGCTGTTTTTGTATACATACAGCAGCAGAGCTGATGCCCTGAAGTATGTCCGATCGAA	2220
Db	2218	GAGCTGTTTTTGTATACATACAGCAGCAGAGCTGATGCCCTGAAGTATGTCCGATCGAA	2277
QY	2221	AGAGAAATGGAATCCCTTGA 2241	
Db	2278	AGAGAAATGGAATCCCTTGA 2298	

RESULT 8  
US-10-342-887-1043  
Sequence 1043, Application US/10342887  
Publication No. US20040058340A1  
GENERAL INFORMATION:  
APPLICANT: Dai, Hongyue  
APPLICANT: He, Yudong  
APPLICANT: Linsley, Peter S.  
APPLICANT: Mao, Mao  
APPLICANT: Robert, Christopher J.  
APPLICANT: Van 't Veer, Laura Johanna  
APPLICANT: Van de Vijver, Marc J.  
APPLICANT: Bernards, Rene  
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
FILE REFERENCE: 9301-188-999  
CURRENT APPLICATION NUMBER: US/10/342,887  
CURRENT FILING DATE: 2003-01-15  
PRIOR APPLICATION NUMBER: 60/298,918  
PRIOR FILING DATE: 2001-06-18  
PRIOR APPLICATION NUMBER: 60/380,710  
PRIOR FILING DATE: 2002-05-14  
PRIOR APPLICATION NUMBER: 10/172,118  
PRIOR FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 2699  
SEQ ID NO 1043  
LENGTH: 2576  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-342-887-1043

Query Match 99.8%; Score 2236.2; DB 8; Length 2576;

Beet Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGCGCAGACTGGAGAAATCTGAGAGGAGCAGTTGTTGGCGGAGCGTGTAAA 60  
DB 58 ATGGGCGCAGACTGGAGAAATCTGAGAGGAGCAGTTGTTGGCGGAGCGTGTAAA 117  
QY 61 TCGAGTACATCGACTGAGACGCTCAAGAGTTCAAGCAGCTGATGAAGTAAAGT 120  
DB 118 TCGAGTACATCGACTGAGACGCTCAAGAGTTCAAGCAGCTGATGAAGTAAAGT 177  
QY 121 ATGTTAGTTCGAATCGTCAAAAATTTTGAAGAAGGAAATCTTAAACGAAGATG 180  
DB 178 ATGTTAGTTCGAATCGTCAAAAATTTTGAAGAAGGAAATCTTAAACGAAGATG 237  
QY 181 AAACAGCGAAGATACAGCTGTGCAATCTGACTTCTGAGCTCATTTGCGCGGACT 240  
DB 238 AAACAGCGAAGATACAGCTGTGCAATCTGACTTCTGAGCTCATTTGCGCGGACT 297  
QY 241 AGGAGTGTTCGCTGACAGCTGACATCTGACTTCTGAGCTCATTTGCGCGGACT 300  
DB 298 AGGAGTGTTCGCTGACAGCTGACATCTGACTTCTGAGCTCATTTGCGCGGACT 357  
QY 301 CTGAATGCACTGCTTCACTACCAATTAATTAATTTCTGCTCCCTTACAGAGATTT 360  
DB 358 CTGAATGCACTGCTTCACTACCAATTAATTAATTTCTGCTCCCTTACAGAGATTT 417  
QY 361 ATGTGGAAGATGAATCTTTTAACTTAATTAATTTCTGCTCCCTTACAGAGATTT 420  
DB 418 ATGTGGAAGATGAATCTTTTAACTTAATTAATTTCTGCTCCCTTACAGAGATTT 477  
QY 421 CAGATGCTACTTCAATGAAGAACTAATTAATTAATTTATGATGGAAGTACACGGGAT 480  
DB 478 CAGATGCTACTTCAATGAAGAACTAATTAATTAATTTATGATGGAAGTACACGGGAT 537  
QY 481 AGAGATGTGGGTATTAATAAGATGAATTTTGTGGAGTTGCTGAATGCTTGTGCTAA 540  
DB 538 AGAGATGTGGGTATTAATAAGATGAATTTTGTGGAGTTGCTGAATGCTTGTGCTAA 597  
QY 541 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
DB 598 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657

QY 601 AAAGATCTGGAGATCAACGAGATGATTAAGAAGCCGCCACTTCGAAATTTCTCTCT 660  
DB 658 AAAGATCTGGAGATCAACGAGATGATTAAGAAGCCGCCACTTCGAAATTTCTCTCT 717  
QY 661 GATTAATTTTGAAGCATTTCTCAATGTTTCAAGATTAAGGACAGCAGAGAACTA 720  
DB 718 GATTAATTTTGAAGCATTTCTCAATGTTTCAAGATTAAGGACAGCAGAGAACTA 777  
QY 721 AAGGAAAAATTAAGAATCAACGAGATGATTAAGAAGCCGCCACTTCGAAATTT 780  
DB 778 AAGGAAAAATTAAGAATCAACGAGATGATTAAGAAGCCGCCACTTCGAAATTT 837  
QY 781 ACCCCCAATGAGAGCAAAATGCTAAATGTTTCAAGAGAGCAAAAGCTTACATCC 840  
DB 838 ACCCCCAATGAGAGCAAAATGCTAAATGTTTCAAGAGAGCAAAAGCTTACATCC 897  
QY 841 TTTCAATGCTTTTCTGATGAGCATGTTTAAATGATGATGATGATGATGATGATGAT 900  
DB 898 TTTCAATGCTTTTCTGATGAGCATGTTTAAATGATGATGATGATGATGATGATGAT 957  
QY 901 GCAACACCAACATTAATTAAGGAGAAACAGAAACAGCTTGAACAACAACTTGT 960  
DB 958 GCAACACCAACATTAATTAAGGAGAAACAGAAACAGCTTGAACAACAACTTGT 1017  
QY 961 GGAACCAAGTGTACAGACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 1018 GGAACCAAGTGTACAGACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077  
QY 1021 GAGCGATTAAGAAGCCCAACAAAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1078 GAGCGATTAAGAAGCCCAACAAAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
QY 1081 AAGAGTACAGAGCCCAACAGCCCAACATTAATGCTGGAATCAAGAGATACAGACT 1140  
DB 1138 AAGAGTACAGAGCCCAACAGCCCAACATTAATGCTGGAATCAAGAGATACAGACT 1197  
QY 1141 GATAGGAG 1200  
DB 1198 GATAGGAG 1257  
QY 1201 AAAGATGAATCTTCAAGCTCTCTGAACAAATTTCTGAGTCAACCAATTAAGATG 1260  
DB 1258 AAAGATGAATCTTCAAGCTCTCTGAACAAATTTCTGAGTCAACCAATTAAGATG 1317  
QY 1261 AAGCAATTAATTAAGCTCTCTGAACAAATTTCTGAGTCAACCAATTAAGATG 1320  
DB 1318 AAGCAATTAATTAAGCTCTCTGAACAAATTTCTGAGTCAACCAATTAAGATG 1377  
QY 1321 AGAGTCTCATTTGAGCATTAATTAAGCAATTTCTGAGTCAACCAATTAAGATG 1380  
DB 1378 AGAGTCTCATTTGAGCATTAATTAAGCAATTTCTGAGTCAACCAATTAAGATG 1437  
QY 1381 AAAACATGATGAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1440  
DB 1438 AAAACATGATGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1497  
QY 1441 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
DB 1498 CCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557  
QY 1501 GCACTGCAAGAAAGATACAGCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
DB 1558 GCACTGCAAGAAAGATACAGCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
QY 1561 CAACCTGATGATCAACAGGAGCTTGTGAAGATGATGATGATGATGATGATGATGAT 1620  
DB 1618 CAACCTGATGATCAACAGGAGCTTGTGAAGATGATGATGATGATGATGATGATGAT 1677  
QY 1621 AATTTTGTGAAGATTTTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 1680  
DB 1678 AATTTTGTGAAGATTTTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 1737



QY 1681 CGCTGCAAGCAGTGCACACCAAGCAGTCCCGTCTACCTGGCTGTCCGAGAGTGT 1740  
|  
|  
|  
Db 1738 CGCTGCAAGCAGTGCACACCAAGCAGTCCCGTCTACCTGGCTGTCCGAGAGTGT 1797  
|  
|  
|  
QY 1741 GACCTGACCTCTGTCTTACTTTGTGTGAGCCCGCTGACCTTTGGGACAGATAAAATGTGTCC 1800  
|  
|  
|  
Db 1798 GACCTGACCTCTGTCTTACTTTGTGTGAGCCCGCTGACCTTTGGGACAGATAAAATGTGTCC 1857  
|  
|  
|  
QY 1801 TGCAGAACTGCAATATTCAGCGGGGCTCCAAAAGCATCTATTGTGCGCACCATCTGAC 1860  
|  
|  
|  
Db 1858 TGCAGAACTGCAATATTCAGCGGGGCTCCAAAAGCATCTATTGTGCGCACCATCTGAC 1917  
|  
|  
|  
QY 1861 GTGCGAGGCTGGGGATTTTATCAAGATCTGTGCAAAAATGAATTCACTCAGAA 1920  
|  
|  
|  
Db 1918 GTGCGAGGCTGGGGATTTTATCAAGATCTGTGCAAAAATGAATTCACTCAGAA 1977  
|  
|  
|  
QY 1921 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAAAGGAAAGTTATGATTA 1980  
|  
|  
|  
Db 1978 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAAAGGAAAGTTATGATTA 2037  
|  
|  
|  
QY 1981 TACATGTGACGCTTCTGTCAACTTGAACATGATTTTGTGTGATGCAACCGCAG 2040  
|  
|  
|  
Db 2038 TACATGTGACGCTTCTGTCAACTTGAACATGATTTTGTGTGATGCAACCGCAG 2097  
|  
|  
|  
QY 2041 GGTACAAAAATTCGTTTTCGTAATCTCGTAAATCCAAACTGCTATGCAAAAAGTTATG 2100  
|  
|  
|  
Db 2098 GGTACAAAAATTCGTTTTCGTAATCTCGTAAATCCAAACTGCTATGCAAAAAGTTATG 2157  
|  
|  
|  
QY 2101 ATGTGTAAACGATGATCAAGATAGGTATTTTGTGCAAGAGAGCATCAAGTGGGAA 2160  
|  
|  
|  
Db 2158 ATGTGTAAACGATGATCAAGATAGGTATTTTGTGCAAGAGAGCATCAAGTGGGAA 2217  
|  
|  
|  
QY 2161 GAGCTGTTTTTGTATTACAGATACAGCCAGGCTGATGCTGAAATATGTGCGCATCGAA 2220  
|  
|  
|  
Db 2218 GAGCTGTTTTTGTATTACAGATACAGCCAGGCTGATGCTGAAATATGTGCGCATCGAA 2277  
|  
|  
|  
QY 2221 AGAGAAATGAAATCCCTTGA 2241  
|  
|  
|  
Db 2278 AGAGAAATGAAATCCCTTGA 2298  
|  
|  
|

RESULT 9  
US-10-848-755A-129  
; Sequence 129, Application US/10848755A  
; Publication No. US20050054826A1  
; GENERAL INFORMATION:  
; APPLICANT: Mao, Mao  
; TITLE OF INVENTION: HUMAN DIAPHANOUS-3 GENE AND METHODS OF USE THEREFOR  
; FILE REFERENCE: 9301-196-999  
; CURRENT APPLICATION NUMBER: US/10/848,755A  
; CURRENT FILING DATE: 2004-05-18  
; PRIOR APPLICATION NUMBER: 60/471,842  
; PRIOR FILING DATE: 2003-05-11  
; NUMBER OF SEQ ID NOS: 275  
; SOFTWARE: PatentIn version 3.2 CAM: 301891-999188  
; SEQ ID NO 129  
; LENGTH: 2576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-848-755A-129

Query Match 99.8%; Score 2236.2; DB 10; Length 2576;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGCGCAGATCTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAAGGTGTAA 60  
|  
|  
|  
Db 58 ATGGGCGCAGATCTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAAGGTGTAA 117  
|  
|  
|  
QY 61 TCAGAGTACATGCGATGAGACAGCTCAAGAGATTCAAGCAGATGATGAATAAGAGT 120  
|  
|  
|  
Db 118 TCAGAGTACATGCGATGAGACAGCTCAAGAGATTCAAGCAGATGATGAATAAGAGT 177  
|  
|  
|

QY 121 ATGTTTAGTTCATATGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATG 180  
|  
|  
|  
Db 178 ATGTTTAGTTCATATGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATG 237  
|  
|  
|  
QY 181 AAACAGGAGAGATACAGCTGTGACATCTGATCTTGTGAGTCAATTCGCGGAGCT 240  
|  
|  
|  
Db 238 AAACAGGAGAGATACAGCTGTGACATCTGATCTTGTGAGTCAATTCGCGGAGCT 297  
|  
|  
|  
QY 241 AGGAGTGTTCGCTGACCAAGTACTTGGATTTTTCACACCAAGTATCCATTAAGACT 300  
|  
|  
|  
Db 298 AGGAGTGTTCGCTGACCAAGTACTTGGATTTTTCACACCAAGTATCCATTAAGACT 357  
|  
|  
|  
QY 301 CTGAATGACATGTGCTTACATACCAATTAATTTTGTGTGCTCCCTACAGCAGAAATTT 360  
|  
|  
|  
Db 358 CTGAATGACATGTGCTTACATACCAATTAATTTTGTGTGCTCCCTACAGCAGAAATTT 417  
|  
|  
|  
QY 361 ATGCTGAGAGATGAACTGTTTTTACATTAATCTTATATGAGGAGATGAGTTTATGAT 420  
|  
|  
|  
Db 418 ATGCTGAGAGATGAACTGTTTTTACATTAATCTTATATGAGGAGATGAGTTTATGAT 477  
|  
|  
|  
QY 421 CAGATGTGATCTTTCATTTGAAGAACTAATATAAAATTTATGATGAGAAAGTACAGGGAT 480  
|  
|  
|  
Db 478 CAGATGTGATCTTTCATTTGAAGAACTAATATAAAATTTATGATGAGAAAGTACAGGGAT 537  
|  
|  
|  
QY 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGTAATGCTTGTGTCA 540  
|  
|  
|  
Db 538 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGTAATGCTTGTGTCA 597  
|  
|  
|  
QY 541 TATATGATGATGACATGATGATGATGAGAGATCTGTAAGAAAGAGAAAGAAAGCAG 600  
|  
|  
|  
Db 598 TATATGATGATGACATGATGATGATGAGAGATCTGTAAGAAAGAGAAAGAAAGCAG 657  
|  
|  
|  
QY 601 AAAGATCTGAGATGACCAAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 660  
|  
|  
|  
Db 658 AAAGATCTGAGATGACCAAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 717  
|  
|  
|  
QY 661 GATTAATTTTGTGAAGCATTCCTCTCAATGTTTCAATTAAGGACACAGCAAGAAACTA 720  
|  
|  
|  
Db 718 GATTAATTTTGTGAAGCATTCCTCTCAATGTTTCAATTAAGGACACAGCAAGAAACTA 777  
|  
|  
|  
QY 721 AAGGAAATATTAAGAACTCACCGAACAGAGCTCCAGGCGCATTCCTCTGAATGT 780  
|  
|  
|  
Db 778 AAGGAAATATTAAGAACTCACCGAACAGAGCTCCAGGCGCATTCCTCTGAATGT 837  
|  
|  
|  
QY 781 ACCCCCAATAGATGACCAATGTCTAATCTGTTCAAGAGACCAAGCTTACACTCC 840  
|  
|  
|  
Db 838 ACCCCCAATAGATGACCAATGTCTAATCTGTTCAAGAGACCAAGCTTACACTCC 897  
|  
|  
|  
QY 841 TTTGATAGCTTTTCTGTAGGCGATGTTTAAATAGACTGCTTCTTAATCTTTCAT 900  
|  
|  
|  
Db 898 TTTGATAGCTTTTCTGTAGGCGATGTTTAAATAGACTGCTTCTTAATCTTTCAT 957  
|  
|  
|  
QY 901 GCAACACCCCAACTTAAGCGGAGAAACACAGAAACAGCTTACAGCAACAACTTGT 960  
|  
|  
|  
Db 958 GCAACACCCCAACTTAAGCGGAGAAACACAGAAACAGCTTACAGCAACAACTTGT 1017  
|  
|  
|  
QY 961 GCAACACAGTGTACAGCAATTTGAGGAGCAAGAGATTGCTGCTCTACCGCT 1020  
|  
|  
|  
Db 1018 GCAACACAGTGTACAGCAATTTGAGGAGCAAGAGATTGCTGCTCTACCGCT 1077  
|  
|  
|  
QY 1021 GAGGGAATTAAGACCCCAACCAAGTCCAGGAGCGCGAAGAGCGGCTTCCCAAT 1080  
|  
|  
|  
Db 1078 GAGGGAATTAAGACCCCAACCAAGTCCAGGAGCGCGAAGAGCGGCTTCCCAAT 1137  
|  
|  
|  
QY 1081 AACAGTACAGAGCCCAAGACCCCAACTTAATGTGCTGAAATCAAGATACAGACGT 1140  
|  
|  
|  
Db 1138 AACAGTACAGAGCCCAAGACCCCAACTTAATGTGCTGAAATCAAGATACAGACGT 1197  
|  
|  
|  
QY 1141 GATAGGAGAGAGGAGCTGAATCGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1200  
|  
|  
|  
Db 1198 GATAGGAGAGAGGAGCTGAATCGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1257  
|  
|  
|  
QY 1201 AAAGATGAATCTTGAGAGTCTCTGAGAGCAAAATTCGGGTGTCAAAACCAATTAAGATG 1260  
|  
|  
|



1258 AAGTGAAGTCTGAGTCTCTGAGAGAAATCTCGGTGCAAAACCATTAAGATG 1317  
1261 AAGCAAAATATGAACTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1320  
1318 AAGCAAAATATGAACTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1377  
1321 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1380  
1378 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1437  
1381 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1440  
1438 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1497  
1441 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1500  
1498 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1557  
1501 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1560  
1558 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1617  
1561 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1620  
1618 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1677  
1621 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1680  
1678 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1737  
1681 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1740  
1738 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1797  
1741 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1800  
1798 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1857  
1801 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1860  
1858 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1917  
1861 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1920  
1918 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1977  
1921 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 1980  
1978 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2037  
1981 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2040  
2038 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2097  
2041 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2100  
2098 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2157  
2101 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2160  
2158 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2217  
2161 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2220  
2218 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2277  
2221 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2280  
2278 AAGTCTCATATGGAAGTCTGAGAAATGAGTGAAGTGTGCTGAAGCTCAATGTTT 2330

RESULT 10

US-10-504-173-78  
; Sequence 78, Application US/10504173  
; Publication No. US20050202428A1  
; GENERAL INFORMATION:  
; APPLICANT: Axordia Limited  
; TITLE OF INVENTION: Pluripotential Stem Cells  
; FILE REFERENCE: P101863WO  
; CURRENT APPLICATION NUMBER: US/10/504,173  
; PRIOR FILING DATE: 2004-08-11  
; PRIOR APPLICATION NUMBER: 0203359.5  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: SeqID version 3.1  
; SEQ ID NO: 78  
; LENGTH: 2576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-504-173-78  
Query Match 99.8%; Score 2236.2; DB 10; Length 2576;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
1 ATGGGCGAGACTGGGAGAGAAATCTGAGAAAGGAGCAAGTTGTTGGCGGAGCTGTAA 60  
58 ATGGGCGAGACTGGGAGAGAAATCTGAGAAAGGAGCAAGTTGTTGGCGGAGCTGTAA 117  
61 TGAAGTATCATGCACTGAGACAGCTCAAGAGTTTCAAGCAAGCTGATGAAGTAAAGT 120  
118 TGAAGTATCATGCACTGAGACAGCTCAAGAGTTTCAAGCAAGCTGATGAAGTAAAGT 177  
121 ATGTTTATGTTTCAATGCTGAGAAATTTTGAAGAAAGGAAATCTTAACCAAGAAATG 180  
178 ATGTTTATGTTTCAATGCTGAGAAATTTTGAAGAAAGGAAATCTTAACCAAGAAATG 237  
181 AAACGAGCAAGATATACAGCTGTGCAATCTGCACTTCTGAGAGTCAATGCGCGGAGT 240  
238 AAACGAGCAAGATATACAGCTGTGCAATCTGCACTTCTGAGAGTCAATGCGCGGAGT 297  
241 AAGGAGTGTGCGTGAAGCAAGTGAATTTTCAACCAAGTCAATCCATTAAGTAAAGT 300  
298 AAGGAGTGTGCGTGAAGCAAGTGAATTTTCAACCAAGTCAATCCATTAAGTAAAGT 357  
301 CTGAATGAGTGTGCTGAGTACCAATTAATGTTTCTGCTGCTGCTGCTGCTGCTGCTG 360  
358 CTGAATGAGTGTGCTGAGTACCAATTAATGTTTCTGCTGCTGCTGCTGCTGCTGCTG 417  
361 ATGTTGGAAGATGAAGCTGTTTCAATTAATGTTTCTGCTGCTGCTGCTGCTGCTGCTG 420  
418 ATGTTGGAAGATGAAGCTGTTTCAATTAATGTTTCTGCTGCTGCTGCTGCTGCTGCTG 477  
421 CAGATGTGATCTTTCATTTGAAGAACTAATAAATAATTAATTAATTAATTAATTAATTA 480  
478 CAGATGTGATCTTTCATTTGAAGAACTAATAAATAATTAATTAATTAATTAATTAATTA 537  
481 AAGAGATGTGTTTAAATGAAGAAATTTTGTGAGTGTGTTGTTGTTGTTGTTGTTGTTG 540  
538 AAGAGATGTGTTTAAATGAAGAAATTTTGTGAGTGTGTTGTTGTTGTTGTTGTTGTTG 597  
541 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
601 AAGATGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
658 AAGATGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
661 GATTAATTTTGAAGCAATTTCTGATGTTTCAATGTTTCAATGTTTCAATGTTTCAATGTT 720  
718 GATTAATTTTGAAGCAATTTCTGATGTTTCAATGTTTCAATGTTTCAATGTTTCAATGTT 777  
721 AAGGAAAAATTAAGAACTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 780

Db 778 AAGGAAAAATATAAGAACTCACCGAACAGAGCTCCAGGCGCATTCCTCTGAAATG 837  
Qy 781 ACCCCCAACATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACACTCC 840  
Db 838 ACCCCCAACATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACACTCC 897  
Qy 841 TTTGATAGGCTTTTCTGTAGGCGAATGTTTAAATATGACTGCTCTTCAATCTTTTCAT 900  
Db 898 TTTGATAGGCTTTTCTGTAGGCGAATGTTTAAATATGACTGCTCTTCAATCTTTTCAT 957  
Qy 901 GCAACACCCCAACATTAAGCGAAGAAACAAGAAAGCTCTAGACAACAAACCTTGT 960  
Db 958 GCAACACCCCAACATTAAGCGAAGAAACAAGAAAGCTCTAGACAACAAACCTTGT 1017  
Qy 961 GGACCAACAGTGTACAGACTTTTGGAGGAGCAAAAGATTGCTGCTGTCAACGCT 1020  
Db 1018 GGACCAACAGTGTACAGACTTTTGGAGGAGCAAAAGATTGCTGCTGTCAACGCT 1077  
Qy 1021 GAGCGGATTAAGACCCCAACAAAGCTCAGAGAGCGCAGAGAAGACGCGCTTCCAT 1080  
Db 1078 GAGCGGATTAAGACCCCAACAAAGCTCAGAGAGCGCAGAGAAGACGCGCTTCCAT 1137  
Qy 1081 AACAGTAGAGGCGCCAGACACCCCACTTAATGTGCTGGAATCAAAAGATATCAGACGT 1140  
Db 1138 AACAGTAGAGGCGCCAGACACCCCACTTAATGTGCTGGAATCAAAAGATATCAGACGT 1197  
Qy 1141 GATAGGGAAGAGGAGCTGAACCGGGGGAGAGAACATGATTAAGAGAGAGAGAG 1200  
Db 1198 GATAGGGAAGAGGAGCTGAACCGGGGGAGAGAACATGATTAAGAGAGAGAGAG 1257  
Qy 1201 AAAGATGAACCTTCAGAGCTCTCTGAGCAAAATTCGCGTGTCAAAACAACATTAAGATG 1260  
Db 1258 AAAGATGAACCTTCAGAGCTCTCTGAGCAAAATTCGCGTGTCAAAACAACATTAAGATG 1317  
Qy 1261 AAGCCAAATATTGAACTCTCTGAGAAATGTGAGATGTGATGTGCTGAAGCCCTCAATGTTT 1320  
Db 1318 AAGCCAAATATTGAACTCTCTGAGAAATGTGAGATGTGATGTGCTGAAGCCCTCAATGTTT 1377  
Qy 1321 AGAGTCTCATTTGGCACTTAATATGACAAATTTCTGTCGACTTGTAGGTTAATGGAGC 1380  
Db 1378 AGAGTCTCATTTGGCACTTAATATGACAAATTTCTGTCGACTTGTAGGTTAATGGAGC 1437  
Qy 1381 AAAACATGTAGACAGGTGTATGATTTAGAGTCAAAAGATCTAGCATATAGCTCAGCT 1440  
Db 1438 AAAACATGTAGACAGGTGTATGATTTAGAGTCAAAAGATCTAGCATATAGCTCAGCT 1497  
Qy 1441 CCGCGTAGAGATGTGATATCTCTCAAGAGAAAGAAAGAGAAACACCGGTGTGTGGCT 1500  
Db 1498 CCGCGTAGAGATGTGATATCTCTCAAGAGAAAGAAAGAGAAACACCGGTGTGTGGCT 1557  
Qy 1501 GCACACTGCAAGAAAGATACAGCTGAAAAAGACGCGCTCCTTAACATATTTTCAACTAT 1560  
Db 1558 GCACACTGCAAGAAAGATACAGCTGAAAAAGACGCGCTCCTTAACATATTTTCAACTAT 1617  
Qy 1561 CAACCTGTATCATCCACGAGGCTTGTGACAGTTGTCGTCGTCCTTGTGTATGACAA 1620  
Db 1618 CAACCTGTATCATCCACGAGGCTTGTGACAGTTGTCGTCGTCCTTGTGTATGACAA 1677  
Qy 1621 AATTTTGTGAAAAATTTTGTCAATGTAGTTCAAGGTGTCAAAACCGCTTTCGGGAATG 1680  
Db 1678 AATTTTGTGAAAAATTTTGTCAATGTAGTTCAAGGTGTCAAAACCGCTTTCGGGAATG 1737  
Qy 1681 CGCTGCAAGCAAGTGTGACACAGAGAGTCCCGGTGTACTGCTGCTGTCCGAGAGT 1740  
Db 1738 CGCTGCAAGCAAGTGTGACACAGAGAGTCCCGGTGTACTGCTGCTGTCCGAGAGT 1797  
Qy 1741 GACCTGTAGCTTGTCTTAATTTGTGAGCGCGTGAACAATGGAACAGTAAAAATGTGTC 1800  
Db 1798 GACCTGTAGCTTGTCTTAATTTGTGAGCGCGTGAACAATGGAACAGTAAAAATGTGTC 1857  
Qy 1801 TGCAAGAACTGAGATTTCAAGGGGGCTCCAAAAAGCATTTATGCTGCAACCATCTAC 1860  
Db 1858 TGCAAGAACTGAGATTTCAAGGGGGCTCCAAAAAGCATTTATGCTGCAACCATCTAC 1917

Qy 1861 GTGGCAGGCTGTGGGGAATTTTATCAAAAGATCTGTGAGAGAAAAATGAATTCATCTAGAA 1920  
Db 1918 GTGGCAGGCTGTGGGGAATTTTATCAAAAGATCTGTGAGAGAAAAATGAATTCATCTAGAA 1977  
Qy 1921 TACTGTGAGAGATTAATTTTCAAGAGTGAAGCTGACAGAAAGAGGAAAGTATGATTA 1980  
Db 1978 TACTGTGAGAGATTAATTTTCAAGAGTGAAGCTGACAGAAAGAGGAAAGTATGATTA 2037  
Qy 1981 TACATGTGACGCTTCTGTCAACTTGAACAATGATTTTGTGTGATGACAAACCGCAG 2040  
Db 2038 TACATGTGACGCTTCTGTCAACTTGAACAATGATTTTGTGTGATGACAAACCGCAG 2097  
Qy 2041 GGTACAAAAATGCTTTTGAACAATCATTCGTAATCAAACTGCTATGCAAAAGTTATG 2100  
Db 2098 GGTACAAAAATGCTTTTGAACAATCATTCGTAATCAAACTGCTATGCAAAAGTTATG 2157  
Qy 2101 ATGCTTAACGATATCAAGATATGATTTTGGCAAGAGACCATCCAGACTGGGAA 2160  
Db 2158 ATGCTTAACGATATCAAGATATGATTTTGGCAAGAGACCATCCAGACTGGGAA 2217  
Qy 2161 GAGCTGTTTTTGTATACAGATACAGCAGGCTGATCCCTGAAGTATGTGCGCATGAA 2220  
Db 2218 GAGCTGTTTTTGTATACAGATACAGCAGGCTGATCCCTGAAGTATGTGCGCATGAA 2277  
Qy 2221 AGAGAAATGGAATCCCTTGA 2241  
Db 2278 AGAGAAATGGAATCCCTTGA 2298

RESULT 11  
US-10-960-414-396  
; Sequence 396, Application US/10960414  
; Publication No. US20060074565A1  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, LANCE D.  
; APPLICANT: GEORGE, JOSIE  
; TITLE OF INVENTION: METHODS, SYSTEMS, AND COMPOSITIONS FOR CLASSIFICATION,  
; TITLE OF INVENTION: METHODS, SYSTEMS, AND COMPOSITIONS FOR CLASSIFICATION,  
; FILE REFERENCE: 38271-76067  
; CURRENT APPLICATION NUMBER: US/10/960,414  
; CURRENT FILING DATE: 2004-10-06  
; NUMBER OF SEQ ID NOS: 500  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 396  
; LENGTH: 2576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-960-414-396

Query Match 99.8%; Score 2236.2; DB 12; Length 2576;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGGGCCAGACTGGAAGAAATCTGAAGAGGACCAAGTTTGTGGCGAAGCGTGTAAA 60  
Db 58 ATGGGCCAGACTGGAAGAAATCTGAAGAGGACCAAGTTTGTGGCGAAGCGTGTAAA 117  
Qy 61 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAAGAGGATGAGTAAAGAT 120  
Db 118 TCAGAGTACATGCACTGAGACAGCTCAAGAGTTTCAAGAGGATGAGTAAAGAT 177  
Qy 121 ATGTTAGTTCATATCGTCAGAAAATTTTGGAAAGAACGAAATCTTAAACCAAGATG 180  
Db 178 ATGTTAGTTCATATCGTCAGAAAATTTTGGAAAGAACGAAATCTTAAACCAAGATG 237  
Qy 181 AAACAGCAGAGATACAGCTGTGCAACTCTGACTTGTGAGCTATTCGCGGAGCT 240  
Db 238 AAACAGCAGAGATACAGCTGTGCAACTCTGACTTGTGAGCTATTCGCGGAGCT 297  
Qy 241 AGGAGCTTGGGAGCAAGTGAATTTTCAACAAGCATTCCTCAATTAAGACT 300

Db 298 AGGAGTGTTCGGTGCACGATGACTGATTTTCCAAACAAGTCATCCATTAAAGACT 357  
Qy 301 CTGAATGCACTGCTTCACTACCCATTAATGTAATCTTGGTCTCCCTTACAGCAATTTT 360  
Db 358 CTGAATGCACTGCTTCACTACCCATTAATGTAATCTTGGTCTCCCTTACAGCAATTTT 417  
Qy 361 ATGTGGAAGATGAATCTGTTTACATTAATCTTATGAGAGATGAAGTTTATGAT 420  
Db 418 ATGTGGAAGATGAATCTGTTTACATTAATCTTATGAGAGATGAAGTTTATGAT 477  
Qy 421 CAGATGCTACTTCACTTGAAGAACTAATTAATTAATGATGGAAGTACACGGAGAT 480  
Db 478 CAGATGCTACTTCACTTGAAGAACTAATTAATTAATGATGGAAGTACACGGAGAT 537  
Qy 481 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGATGTGATATCCCTTGTCA 540  
Db 538 AGAGATGTGGTTTATTAATGATGAATTTTGTGAGATGTGATATCCCTTGTCA 597  
Qy 541 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 598 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
Qy 601 AAAATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCCACTCGAAATTTCTCT 660  
Db 658 AAAATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCCACTCGAAATTTCTCT 717  
Qy 661 GATTAATTTTGAAGCCATTTCTCATGTTTCCAGATTAAGGCAACGAGAAATCTA 720  
Db 718 GATTAATTTTGAAGCCATTTCTCATGTTTCCAGATTAAGGCAACGAGAAATCTA 777  
Qy 721 AAGAAAAATATAAGAACTCACCGAACGAGCTCCGAGGCACTTCTCTGATATG 780  
Db 778 AAGAAAAATATAAGAACTCACCGAACGAGCTCCGAGGCACTTCTCTGATATG 837  
Qy 781 ACCCCCAATAGATGACCAATGCTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 840  
Db 838 ACCCCCAATAGATGACCAATGCTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 897  
Qy 841 TTTCAATGCTTTTCTGTAAGCGATGTTTAAATATGACTGCTTCTCATCTTTTCA 900  
Db 898 TTTCAATGCTTTTCTGTAAGCGATGTTTAAATATGACTGCTTCTCATCTTTTCA 957  
Qy 901 GCAACACCAACACTTATTAAGGGAAGAAACAGAAACAGCTCTGACAAACCTTGT 960  
Db 958 GCAACACCAACACTTATTAAGGGAAGAAACAGAAACAGCTCTGACAAACCTTGT 1017  
Qy 961 GGACCAAGTGTTCACAGATTTGAGAGGAGCAAAAGTGTGCTGCTCTCAACGCT 1020  
Db 1018 GGACCAAGTGTTCACAGATTTGAGAGGAGCAAAAGTGTGCTGCTCTCAACGCT 1077  
Qy 1021 GAGCGATTAAGACCCCAACAAACGTTCCAGAGGCGGCAAGAGACGGCTTCCAT 1080  
Db 1078 GAGCGATTAAGACCCCAACAAACGTTCCAGAGGCGGCAAGAGACGGCTTCCAT 1137  
Qy 1081 AACATGAGAGGCGGCAACCCCAACATTAATGAGTGAATCAAAAGATCAACAAGT 1140  
Db 1138 AACATGAGAGGCGGCAACCCCAACATTAATGAGTGAATCAAAAGATCAACAAGT 1197  
Qy 1141 GATAGGGAAGCAGGATCTGAACCGGGGAGAGAACTAATGATTAAGAAAGAGAGAG 1200  
Db 1198 GATAGGGAAGCAGGATCTGAACCGGGGAGAGAACTAATGATTAAGAAAGAGAGAG 1257  
Qy 1201 AAAGATGAACCTTCAAGCTCTCTGAAGCAAAATCTCGGTCAAAACCAATTAAGATG 1260  
Db 1258 AAAGATGAACCTTCAAGCTCTCTGAAGCAAAATCTCGGTCAAAACCAATTAAGATG 1317  
Qy 1261 AAGCAATATTAAGACCTCTGAAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1320  
Db 1318 AAGCAATATTAAGACCTCTGAAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1377  
Qy 1321 AGAGTCTCATTTGCACTTATGATCAATTTCTGTCGCAATGCTAGGTTAATGGGACC 1380  
Db 1378 AGAGTCTCATTTGCACTTATGATCAATTTCTGTCGCAATGCTAGGTTAATGGGACC 1437

Qy 1381 AAAACATGTAAGCAGGTGTAATGATTTAAGATCAAAAGATCTAGCATCATGCTCAGCT 1440  
Db 1438 AAAACATGTAAGCAGGTGTAATGATTTAAGATCAAAAGATCTAGCATCATGCTCAGCT 1497  
Qy 1441 CCCGCTGAGATGTAATCTCTTCAAGAAAAAGAGAGAAACACGGTGTGGCT 1500  
Db 1498 CCCGCTGAGATGTAATCTCTTCAAGAAAAAGAGAGAAACACGGTGTGGCT 1557  
Qy 1501 GCACTGCAAGAAATACAGCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
Db 1558 GCACTGCAAGAAATACAGCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
Qy 1561 CAACCTGTGATCAACCAAGGAGGCTGTGACAGTTGTGCTGTGTGTGTGTGTGTGT 1620  
Db 1618 CAACCTGTGATCAACCAAGGAGGCTGTGACAGTTGTGCTGTGTGTGTGTGTGTGT 1677  
Qy 1621 AATTTTGTGAAGATTTTGTCAATGTAGTTCAAGATGTCAAAACCGCTTTCGGAGATG 1680  
Db 1678 AATTTTGTGAAGATTTTGTCAATGTAGTTCAAGATGTCAAAACCGCTTTCGGAGATG 1737  
Qy 1681 CGCTGCAAGACAGTGAACACCAAGAGTGCCTGCTACCTGCTGTCCGAGAGTGT 1740  
Db 1738 CGCTGCAAGACAGTGAACACCAAGAGTGCCTGCTACCTGCTGTCCGAGAGTGT 1797  
Qy 1741 GACCTGACCTCTGTCTTACTTGTGTGAGAGCCGCTGACCAATGGAACATGTGTCC 1800  
Db 1798 GACCTGACCTCTGTCTTACTTGTGTGAGAGCCGCTGACCAATGGAACATGTGTCC 1857  
Qy 1801 TGCAAGACTGATGATTCAGAGGAGGCTCAAAAGATCTTATGCTGACCATCTGAC 1860  
Db 1858 TGCAAGACTGATGATTCAGAGGAGGCTCAAAAGATCTTATGCTGACCATCTGAC 1917  
Qy 1861 GTGGCAGGCTGGGAGATTTTATCAAAAGATCTGTGCAAAAGATCAATCTCAGAA 1920  
Db 1918 GTGGCAGGCTGGGAGATTTTATCAAAAGATCTGTGCAAAAGATCAATCTCAGAA 1977  
Qy 1921 TACTGTGAGAGATTTATTTCTCAAGATGATGATGATGATGATGATGATGATGAT 1980  
Db 1978 TACTGTGAGAGATTTATTTCTCAAGATGATGATGATGATGATGATGATGATGAT 2037  
Qy 1981 TACATGTCAGCTTCTGTCAACTTGAACATGATTTTGTGAGATGACCAACCGCAG 2040  
Db 2038 TACATGTCAGCTTCTGTCAACTTGAACATGATTTTGTGAGATGACCAACCGCAG 2097  
Qy 2041 GGTAAACAAATTCGTTTGAACATCTTGTGTAATTCGTAATCTGATGCAAAAGTATG 2100  
Db 2098 GGTAAACAAATTCGTTTGAACATCTTGTGTAATTCGTAATCTGATGCAAAAGTATG 2157  
Qy 2101 ATGTGTAACGGTGAATCAAGATGATGATTTTGTGCAAGAGACCATCAAGTGGCAA 2160  
Db 2158 ATGTGTAACGGTGAATCAAGATGATGATTTTGTGCAAGAGACCATCAAGTGGCAA 2217  
Qy 2161 GAGCTGTTTTTGAATTAAGATTAAGCAGGCTGATGCTGAAGTATGTGGCATGAA 2220  
Db 2218 GAGCTGTTTTTGAATTAAGATTAAGCAGGCTGATGCTGAAGTATGTGGCATGAA 2277  
Qy 2221 AGAGAAATGAAATCCCTTGA 2241  
Db 2278 AGAGAAATGAAATCCCTTGA 2298

RESULT 12  
US-10-210-120-95  
; Sequence 95, Application US/10210120  
; Publication No. US2003017536A1  
; GENERAL INFORMATION:  
; APPLICANT: Chimalyan, Arul M.  
; APPLICANT: Rubin, Mark A.  
; APPLICANT: Steekumar, Arun  
; TITLE OF INVENTION: Expression Profile of Prostate Cancer  
; FILE REFERENCE: UM-07221  
; CURRENT APPLICATION NUMBER: US/10/210.120

CURRENT FILING DATE: 2002-08-01  
PRIOR APPLICATION NUMBER: US 60/309,581  
PRIOR FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: US 60/334,468  
PRIOR FILING DATE: 2001-11-15  
NUMBER OF SEQ ID NOS: 123  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 95  
LENGTH: 2512  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-210-120-95

Query Match 99.5%; Score 2229; DB 7; Length 2512;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2240; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAAGCTGTAAAA 60  
DB 10 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAAGCTGTAAAA 69  
QY 61 TCAGAGTACATGCGCATGAGACAGCTCAAGAGTTTCAAGAGCTGATGAAGTAAAGT 120  
DB 70 TCAGAGTACATGCGCATGAGACAGCTCAAGAGTTTCAAGAGCTGATGAAGTAAAG-GT 128  
QY 121 ATGTTAGTTCATCGTCAGAAAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 180  
DB 129 ATGTTAGTTCATCGTCAGAAAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 188  
QY 181 AAACGCGAAGATTCAGCTGCTGTGACATCTGTAATTTCTGAGTCAATTGCGGGAGCT 240  
DB 189 AAACGCGAAGATTCAGCTGCTGTGACATCTGTAATTTCTGAGTCAATTGCGGGAGCT 248  
QY 241 AGGAGTGTGGGTGACCAAGTACTGATTTTCCAAACAAGTATCCCAATTAAGACT 300  
DB 249 AGGAGTGTGGGTGACCAAGTACTGATTTTCCAAACAAGTATCCCAATTAAGACT 308  
QY 301 CTGAATGCAATGCTTCAGTACCAATTAATTTCTTGGTCTCCCTTACAGCAAAATTTT 360  
DB 309 CTGAATGCAATGCTTCAGTACCAATTAATTTCTTGGTCTCCCTTACAGCAAAATTTT 368  
QY 361 ATGTGGAAGATGAACTGTTTACATAAATCTTCTATATGAGAGATGAAGTTTAAAT 420  
DB 369 ATGTGGAAGATGAACTGTTTACATAAATCTTCTATATGAGAGATGAAGTTTAAAT 428  
QY 421 CAGATGTGATCTTCAATGAAGAACTAATPAAAAATTAATGAGGAAATGCAACGGAGAT 480  
DB 429 CAGATGTGATCTTCAATGAAGAACTAATPAAAAATTAATGAGGAAATGCAACGGAGAT 488  
QY 481 AGAGATGTGGGTATTAATATGATGAAATTTTGTGAGTTGGTGAATGCCCCTTGGTCAA 540  
DB 489 AGAGATGTGGGTATTAATATGATGAAATTTTGTGAGTTGGTGAATGCCCCTTGGTCAA 548  
QY 541 TATATGATGATGAGATGATGATGATGAGACATCTCTGAAGAAAGAGAAAGAAAGCAG 600  
DB 549 TATATGATGATGAGATGATGATGATGAGACATCTCTGAAGAAAGAGAAAGAAAGCAG 608  
QY 601 AAAAGTCTGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 660  
DB 609 AAAAGTCTGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 668  
QY 661 GATAAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAACAGCAAGAACTA 720  
DB 669 GATAAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAACAGCAAGAACTA 728  
QY 721 AAAGAAAAATTAAGAACTCACCGAACAGCAGCTCCAGGGGCACTTCTCTGAAATGT 780  
DB 729 AAAGAAAAATTAAGAACTCACCGAACAGCAGCTCCAGGGGCACTTCTCTGAAATGT 788  
QY 781 ACCCCCAATATGAGAGCAAAATGCTAAATCTGTTAGAGAGCAAAAGCTTACACTCC 840  
DB 789 ACCCCCAATATGAGAGCAAAATGCTAAATCTGTTAGAGAGCAAAAGCTTACACTCC 848

QY 841 TTTTCATACGCTTTTCTGTAGCCGATGTTTAAATATGACTGCTTCTTCAATCCTTTTCAT 900  
DB 849 TTTTCATACGCTTTTCTGTAGCCGATGTTTAAATATGACTGCTTCTTCAATCCTTTTCAT 908  
QY 901 GCAACACCAACCTTATAGCGGAGAACACAGAAACAGCTCTAGCAACAACTTGT 960  
DB 909 GCAACACCAACCTTATAGCGGAGAACACAGAAACAGCTCTAGCAACAACTTGT 968  
QY 961 GCAACACCTGTTAACAGATTTGAGAGAGCAAAAGAGTTTGTCTGCTTCAACCGCT 1020  
DB 969 GCAACACCTGTTAACAGATTTGAGAGAGCAAAAGAGTTTGTCTGCTTCAACCGCT 1028  
QY 1021 GAGCGATTAAGACCCCAACAAACCTTCCAGAGAGCCGCGAAGAGGACCGCTCCCAAT 1080  
DB 1029 GAGCGATTAAGACCCCAACAAACCTTCCAGAGAGCCGCGAAGAGGACCGCTCCCAAT 1088  
QY 1081 AACAGTACAGGCTCAGACACCCCAACCAATTAATGTCTGAAATCAAGATATACAGACGT 1140  
DB 1089 AACAGTACAGGCTCAGACACCCCAACCAATTAATGTCTGAAATCAAGATATACAGACGT 1148  
QY 1141 GATAGGAGACAGGACTGAAAACGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1200  
DB 1149 GATAGGAGACAGGACTGAAAACGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1208  
QY 1201 AAAATGAAATCTTGAAGCTCTCTGAGCAAAATTTCTGAGTCAAAACCAATTAAGATG 1260  
DB 1209 AAAATGAAATCTTGAAGCTCTCTGAGCAAAATTTCTGAGTCAAAACCAATTAAGATG 1268  
QY 1261 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGAGTGTGTGAGAGCTCAATGTTT 1320  
DB 1269 AAGCCAAATATTTGAACCTCTGAGAAATGTGAGAGTGTGTGAGAGCTCAATGTTT 1328  
QY 1321 AAGTCTCTATTTGACCTTATGACAAATTTCTGTCCATTTGTAGTTAATTTGGAGC 1380  
DB 1329 AAGTCTCTATTTGACCTTATGACAAATTTCTGTCCATTTGTAGTTAATTTGGAGC 1388  
QY 1381 AAAACATGTAGACAGGTGTATGAGTTTAAAGTAAAGTCAATAGCTCAAGCT 1440  
DB 1389 AAAACATGTAGACAGGTGTATGAGTTTAAAGTAAAGTCAATAGCTCAAGCT 1448  
QY 1441 CCGCTGAGATGTGATTAATCTCTCAAGGAAAGAAAGAGAAACACCGGTTGTGGGCT 1500  
DB 1449 CCGCTGAGATGTGATTAATCTCTCAAGGAAAGAAAGAGAAACACCGGTTGTGGGCT 1508  
QY 1501 GCAACCTGCAAGAAAGATACAGCTGAAAGAGACGGCTCTCTTAACATGTTTACACTAT 1560  
DB 1509 GCAACCTGCAAGAAAGATACAGCTGAAAGAGACGGCTCTCTTAACATGTTTACACTAT 1568  
QY 1561 CAACCTGTGATTAATCAACGGCAACCTTGTGACAGTTCTGTGCTGTGTGATGACAA 1620  
DB 1569 CAACCTGTGATTAATCAACGGCAACCTTGTGACAGTTCTGTGCTGTGTGATGACAA 1628  
QY 1621 AATTTTGTGAAAGTTTGTCAATGTAGTTCAAGTGTCAAAACCGCTTCCGGGATGC 1680  
DB 1629 AATTTTGTGAAAGTTTGTCAATGTAGTTCAAGTGTCAAAACCGCTTCCGGGATGC 1688  
QY 1681 CGCTGCAAAAGCAAGTGCACCAACAGAGTCCCTGTCTACTCGTGTCTCGAGAGTGT 1740  
DB 1689 CGCTGCAAAAGCAAGTGCACCAACAGAGTCCCTGTCTACTCGTGTCTCGAGAGTGT 1748  
QY 1741 GACCTTGAACCTTCTTATCTTGTGAGCCGCTGACCAATTTGGGACAGTAAAGTGTCC 1800  
DB 1749 GACCTTGAACCTTCTTATCTTGTGAGCCGCTGACCAATTTGGGACAGTAAAGTGTCC 1808  
QY 1801 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGACATTAATTTGTGCACTTGAC 1860  
DB 1809 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGACATTAATTTGTGCACTTGAC 1868  
QY 1861 GTGCAAGCTGTGGGAGATTTTATCAAAAGATCTGTGCAAGAAATGAAATTCATCTCAGAA 1920  
DB 1869 GTGCAAGCTGTGGGAGATTTTATCAAAAGATCTGTGCAAGAAATGAAATTCATCTCAGAA 1928  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAGTATGATTA 1980

Db 1929 TACTGTGAGAGATTTTCTGAAAGTGAAGGAGGAGGAAAGTATGATTA 1988  
Qy 1981 TACATGTGACGCTTTCTGTTCAATGAAATGATTTGTGTGATGACACCGGCAAG 2040  
Db 1989 TACATGTGACGCTTTCTGTTCAATGAAATGATTTGTGTGATGACACCGGCAAG 2048  
Qy 2041 GGTAAACAAATTCGTTTGCATTCATGCTGATTAATCCAACTGATCAAAAGTTAG 2100  
Db 2049 GGTAAACAAATTCGTTTGCATTCATGCTGATTAATCCAACTGATCAAAAGTTAG 2108  
Qy 2101 ATGTTAAGCGTGTATCAAGATAGGATTTTTCAGAGAGCCATCCAGATGCGCA 2160  
Db 2109 ATGTTAAGCGTGTATCAAGATAGGATTTTTCAGAGAGCCATCCAGATGCGCA 2168  
Qy 2161 GAGCTGTTTTTGTATTAAGATACAGCCAGCGCTGATGCTTGAATGTCCGATCGAA 2220  
Db 2169 GAGCTGTTTTTGTATTAAGATACAGCCAGCGCTGATGCTTGAATGTCCGATCGAA 2228  
Qy 2221 AGAGAAATGAAATCCCTTGA 2241  
Db 2229 AGAGAAATGAAATCCCTTGA 2249

RESULT 13  
US-10-909-035-95  
; Sequence 95, Application US/10909035  
; Publication No. US20050136493A1  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Mark A.  
; APPLICANT: Chinnaiyan, Arul M.  
; APPLICANT: Laxman, Bharathi  
; APPLICANT: Steekumar, Arun  
; TITLE OF INVENTION: AMACR Cancer Markers  
; FILE REFERENCE: UM-09098  
; CURRENT APPLICATION NUMBER: US/10/909, 035  
; CURRENT FILING DATE: 2004-07-30  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 95  
; LENGTH: 2512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-909-035-95

Query Match 99.5%; Score 2229; DB 10; Length 2512;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2240; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGAGCCAGTTTGTGGCGGAAAGCGTGTAA 60  
Db 10 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGAGCCAGTTTGTGGCGGAAAGCGTGTAA 69  
Qy 61 TCAGAGTACATCGACTGAGACAGCTCAAGAGGTTTCAAGAGCTGATGAAGTAAAGT 120  
Db 70 TCAGAGTACATCGACTGAGACAGCTCAAGAGGTTTCAAGAGCTGATGAAGTAAAG-GT 128  
Qy 121 ATGTTAAGTTCCAAATCGTCAAAAATTTTGAAGAAGCGAAATCTTAAACCAAGATG 180  
Db 129 ATGTTAAGTTCCAAATCGTCAAAAATTTTGAAGAAGCGAAATCTTAAACCAAGATG 188  
Qy 181 AAACAGCCAGAAATACAGCTGTGACATCTCTGTGAGCTCATTTGGCGGAGCT 240  
Db 189 AAACAGCCAGAAATACAGCTGTGACATCTCTGTGAGCTCATTTGGCGGAGCT 248  
Qy 241 AGGAGTGTTCGTTGACAGTACCTGATTTTCCAAACAAGTCAATCCATTAAAGCT 300  
Db 249 AGGAGTGTTCGTTGACAGTACCTGATTTTCCAAACAAGTCAATCCATTAAAGCT 308  
Qy 301 CTGAATGAGTTGCTTCACTACATCAATATGATTTCTGTGCTCCCTTACAGAGATTTT 360  
Db 309 CTGAATGAGTTGCTTCACTACATCAATATGATTTCTGTGCTCCCTTACAGAGATTTT 368

Qy 361 ATGTGGAAGATGAAATCTGTTTATCAATACATTCCTTATATGAGAGATGAAGTTTAT 420  
Db 369 ATGTGGAAGATGAAATCTGTTTATCAATACATTCCTTATATGAGAGATGAAGTTTAT 428  
Qy 421 CAGATGTACTCTTCAATGAAAGTAAATATAAATTTATGATGGGAAAGTACACGGGAT 480  
Db 429 CAGATGTACTCTTCAATGAAAGTAAATATAAATTTATGATGGGAAAGTACACGGGAT 488  
Qy 481 AGAGATGTGGGTTTATATATGATGAATTTTGTGAGTGTGTGAATGCCCTGTGCA 540  
Db 489 AGAGATGTGGGTTTATATATGATGAATTTTGTGAGTGTGTGAATGCCCTGTGCA 548  
Qy 541 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 549 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608  
Qy 601 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCACTCGAAATTTCTTCT 660  
Db 609 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCACTCGAAATTTCTTCT 668  
Qy 661 GATTAATTTTGAAGCCATTTCTCATATGTTTCCAGATAGGGCAACGAGAAAGTCA 720  
Db 669 GATTAATTTTGAAGCCATTTCTCATATGTTTCCAGATAGGGCAACGAGAAAGTCA 728  
Qy 721 AAGAAAAATATAAAGAACTCAACGAGAGAGCTCCAGGCGCATTCCTCTGAAATG 780  
Db 729 AAGAAAAATATAAAGAACTCAACGAGAGAGCTCCAGGCGCATTCCTCTGAAATG 788  
Qy 781 ACCCCCAATATGATGAGCAAAATCTAATCTGTTCAAGAGAGCAAAAGTCACTCC 840  
Db 789 ACCCCCAATATGATGAGCAAAATCTAATCTGTTCAAGAGAGCAAAAGTCACTCC 848  
Qy 841 TTTCAATAGCTTTTCTGATAGCGGATTTTAAATATGATGCTTCTTACATCTTTTAT 900  
Db 849 TTTCAATAGCTTTTCTGATAGCGGATTTTAAATATGATGCTTCTTACATCTTTTAT 908  
Qy 901 GCAACACCAACACTATTAAGCGGAGAAACAGAAAGAGCTCTGACAAACAACTTGT 960  
Db 909 GCAACACCAACACTATTAAGCGGAGAAACAGAAAGAGCTCTGACAAACAACTTGT 968  
Qy 961 GGACCAAGTGTATCAAGACTTTTGAAGGAGAGCAAGATTTGTCTGCTCTCAACGCT 1020  
Db 969 GGACCAAGTGTATCAAGACTTTTGAAGGAGAGCAAGATTTGTCTGCTCTCAACGCT 1028  
Qy 1021 GAGCGATTAAGACCCCAACAAACGTCACAGAGCCCGCAAGAGAGAGCGCTTCCAT 1080  
Db 1029 GAGCGATTAAGACCCCAACAAACGTCACAGAGCCCGCAAGAGAGAGCGCTTCCAT 1088  
Qy 1081 AACATACAGGCCCCAGACCCCACTTAATGTGCTGGAATGCAAGATPACAGACT 1140  
Db 1089 AACATACAGGCCCCAGACCCCACTTAATGTGCTGGAATGCAAGATPACAGACT 1148  
Qy 1141 GATAGGAG 1200  
Db 1149 GATAGGAG 1208  
Qy 1201 AAAGATGAATCTTGAGCTCTCTGAGCAAAATCTCGGTCAAAACCAATTAAGATG 1260  
Db 1209 AAAGATGAATCTTGAGCTCTCTGAGCAAAATCTCGGTCAAAACCAATTAAGATG 1268  
Qy 1261 AAAGCAATATTAAGACCTCTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1269 AAAGCAATATTAAGACCTCTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1328  
Qy 1321 AAGATCTCATTTGAGCACTTATGACAAATTTCTGTGCAATTTGATGATTTGGGAC 1380  
Db 1329 AAGATCTCATTTGAGCACTTATGACAAATTTCTGTGCAATTTGATGATTTGGGAC 1388  
Qy 1381 AAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATATGCTCAACT 1440  
Db 1389 AAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATATGCTCAACT 1448  
Qy 1441 CCGGTGAGATGTGATPACTCTCCAAAGAAAAAGAGAGAAACACCGGTTGTGGCT 1500

1449 CCCCCGAGATGAGATACCTCCAGAGAAAAAGAGAGAAACCGGTTGGGCT 1508  
QY GCACATGTCAGAAAAGATACAGCTGAAAAAGACGGCTCTCTAACCATGTTTACAACTAT 1560  
Db GACACATGTCAGAAAAGATACAGCTGAAAAAGACGGCTCTCTAACCATGTTTACAACTAT 1568  
QY 1561 CAACCTGTGATCATCCAGCGGACGCTTGTGACAGTTGTGTCCTTGTGTGATAGCAAA 1620  
Db 1569 CAACCTGTGATCATCCAGCGGACGCTTGTGACAGTTGTGTCCTTGTGTGATAGCAAA 1628  
QY 1621 AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGGTCTCAAAAACCGCTTCCGGAGTC 1680  
Db 1629 AATTTTGTGAAAAAGTTTGTCAATGTATGTTCAAGGTCTCAAAAACCGCTTCCGGAGTC 1688  
QY 1681 CGCTGCAGAAAGCAAGTGCAGAACCAAGAGAGTGCCTGCTCACTGGCTGTCCGAGAGTGT 1740  
Db 1689 CGCTGCAGAAAGCAAGTGCAGAACCAAGAGAGTGCCTGCTCACTGGCTGTCCGAGAGTGT 1748  
QY 1741 GACCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATTGGGACAGTAAAAATGTGTCC 1800  
Db 1749 GACCTGACCTCTGTCTTACTTGTGTGAGCCGCTGACCAATTGGGACAGTAAAAATGTGTCC 1808  
QY 1801 TGCAGAAACTGCAATGATTCAGCGGGGCTCCAAAAAGCATTTTGTCTGGCCATCTTAC 1860  
Db 1809 TGCAGAAACTGCAATGATTCAGCGGGGCTCCAAAAAGCATTTTGTCTGGCCATCTTAC 1868  
QY 1861 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCTGTGCAGAAAAATGAATTCATCTCAGAA 1920  
Db 1869 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCTGTGCAGAAAAATGAATTCATCTCAGAA 1928  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAGAGCTGACAGAGAGGAAAGTGTATGATAA 1980  
Db 1929 TACTGTGAGAGATTAATTTCTCAAGATGAGAGCTGACAGAGAGGAAAGTGTATGATAA 1988  
QY 1981 TACATGTGACATTTCTGTGTTCACTTGAACAATGATTTGTGTGTGAGTGCACCCGCAAG 2040  
Db 1989 TACATGTGACATTTCTGTGTTCACTTGAACAATGATTTGTGTGTGAGTGCACCCGCAAG 2048  
QY 2041 GGTAAACAATAATCGTTTGCAGAAATCAATCGGTAAATCCAAACTGCTATGCAAAAAGTTATG 2100  
Db 2049 GGTAAACAATAATCGTTTGCAGAAATCAATCGGTAAATCCAAACTGCTATGCAAAAAGTTATG 2108  
QY 2101 ATGTTAAACGTTGATCACAGATAGGTATTTTGTCCAAAGAGCCATCCAGCTGGCGAA 2160  
Db 2109 ATGTTAAACGTTGATCACAGATAGGTATTTTGTCCAAAGAGCCATCCAGCTGGCGAA 2168  
QY 2161 GAGCTGTTTTTGTGATTAAGATACAGCCAGGCTGATGCTCCGAAGTATGTGGGCAATCGAA 2220  
Db 2169 GAGCTGTTTTTGTGATTAAGATACAGCCAGGCTGATGCTCCGAAGTATGTGGGCAATCGAA 2228  
QY 2221 AGAGAAATGAGAAATCCCTTGA 2241  
Db 2229 AGAGAAATGAGAAATCCCTTGA 2249

RESULT 14  
US-10-104-047-1192

Sequence 1192, Application US/10104047  
Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
FILE REFERENCE: H1-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1192  
LENGTH: 2476  
TYPE: DNA

ORGANISM: Homo sapiens  
US-10-104-047-1192  
Query Match 89.0%; Score 1993.8; DB 7; Length 2476;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 2; Indels 117; Gaps 1;  
QY 1 ATGGGCCAGACTGGGAGAAATCTGAGAGAGACCAAGTTGTTGGCGGAACCGTGTAAA 60  
Db 93 ATGGGCCAGACTGGGAGAAATCTGAGAGAGACCAAGTTGTTGGCGGAACCGTGTAAA 152  
QY 61 TCAGAGTACATGGGACCTGAGACAGCTCAAGAGGTTGAGAGCTGATGAAAGTAAAGT 120  
Db 153 TCAGAGTACATGGGACCTGAGACAGCTCAAGAGGTTGAGAGCTGATGAAAGTAAAGT 212  
QY 121 ATGTTAGTTCATCTGAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 180  
Db 213 ATGTTAGTTCATCTGAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 272  
QY 181 AAACAGGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGCCGGGACT 240  
Db 273 AAACAGGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGCCGGGACT 332  
QY 241 AGGAGGTGTCGGTGCACAGTGAATTGGAATTTTCCAAACAAGTCATCCATTAAAGACT 300  
Db 333 AGGGA----- 337  
QY 301 CTGAATCAGTGTCTTCAATACCATTAATGATTTCTTGCTCCCTACAGAGAAATTTT 360  
Db 338 ----- 337  
QY 361 ATGTGTGAGATGAACTGTTTATCATTAACATTTCTTATATGGAGATGAAGTTTATGAT 420  
Db 338 --GGTGAAGATGAACTGTTTATCATTAACATTTCTTATATGGAGATGAAGTTTATGAT 395  
QY 421 CAGGATGGTACTTCAATTTGAAAGAACTAATAAATAATATGATGGGAAAGTACAGGGGAT 480  
Db 396 CAGGATGGTACTTCAATTTGAAAGAACTAATAAATAATATGATGGGAAAGTACAGGGGAT 455  
QY 481 AGAGATGTGGGTTTAAATGATGAATTTTGTGAGGTTGATGAAATGCCCTTGTCTCA 540  
Db 456 AGAGATGTGGGTTTAAATGATGAATTTTGTGAGGTTGATGAAATGCCCTTGTCTCA 515  
QY 541 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Db 516 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 575  
QY 601 AAAGATCGAGGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660  
Db 576 AAAGATCGAGGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 635  
QY 661 GATTAATTTTGAAGCATTTCTCTCAATGTTTCCAGTAAAGGCAACGAGAAAGAACTA 720  
Db 636 GATTAATTTTGAAGCATTTCTCTCAATGTTTCCAGTAAAGGCAACGAGAAAGAACTA 695  
QY 721 AAGGAAAAATTAAGAACTACAGGACAGAGCTCCAGGCGCACTTCTCTGATATGT 780  
Db 696 AAGGAAAAATTAAGAACTACAGGACAGAGCTCCAGGCGCACTTCTCTGATATGT 755  
QY 781 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACATCC 840  
Db 756 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACATCC 815  
QY 841 TTTGATACGCTTTTCTGAGCGATGTTTAAATATGATGCTTCTTCAATCCTTTTCAT 900  
Db 816 TTTGATACGCTTTTCTGAGCGATGTTTAAATATGATGCTTCTTCAATCCTTTTCAT 875  
QY 901 GCAACACCCCACTTATATAGCGGAGAGACACAGAAACAGCTCTAGACAAACAACTTGT 960  
Db 876 GCAACACCCCACTTATATAGCGGAGAGACACAGAAACAGCTCTAGACAAACAACTTGT 935  
QY 961 GGACCAAGTGTTCACGACATTTGAGGAGAGCAAAAGGTTTGTGCTGCTCAGCGCT 1020







QY 241 AGGAGTGTTCGGTGACCACTGACTTGATTTTCCACACAAAGTCATCCATTAAGACT 300  
| | | | |  
Db 333 AGGGA----- 337  
QY 301 CTGAATGACGTTGCTTCAGTACCAATATGTAATCTTGCTGCCCTACAGACAAATTTT 360  
| | | | |  
Db 338 ----- 337  
QY 361 ATGTGTGAAGATGAAACTGTTTTCATTAACATTCCTTATATGAGATGAAGTTTATAGT 420  
| | | | |  
Db 338 --GGTGGAAAGATGAACCTGTTTACATACATTCCTTATATGAGAGATGAAGTTTATAGT 395  
| | | | |  
QY 421 CAGAGTGTACTTTTCATTGAAAGAACTAATTAATAAATTATGATGGAAAGTACCGGGAT 480  
| | | | |  
Db 396 CAGAGTGTACTTTTCATTGAAAGAACTAATTAATAAATTATGATGGAAAGTACCGGGAT 455  
| | | | |  
QY 481 AGAAGATGTGGGTTTATTAATGAAGAAATTTTGTGGGTTGGTGAAGCCCTTGCTGAA 540  
| | | | |  
Db 456 AGAAGATGTGGGTTTATTAATGAAGAAATTTTGTGGGTTGGTGAAGCCCTTGCTGAA 515  
| | | | |  
QY 541 TATATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
| | | | |  
Db 516 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 575  
| | | | |  
QY 601 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 660  
| | | | |  
Db 576 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 635  
| | | | |  
QY 661 GATAAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAGAAAGACTA 720  
| | | | |  
Db 636 GATAAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAGAAAGACTA 695  
| | | | |  
QY 721 AAGGAAATTAATAAGAACTCACCGAACAGAGCTCCAGGGCGCACTTCTCTGAAATGT 780  
| | | | |  
Db 696 AAGGAAATTAATAAGAACTCACCGAACAGAGCTCCAGGGCGCACTTCTCTGAAATGT 755  
| | | | |  
QY 781 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
| | | | |  
Db 756 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 815  
| | | | |  
QY 841 TTTCAATAGCTTTCTGTAAGGAGATGTTTAAATATGATGATGATGATGATGATGATGATG 900  
| | | | |  
Db 816 TTTCAATAGCTTTCTGTAAGGAGATGTTTAAATATGATGATGATGATGATGATGATGATG 875  
| | | | |  
QY 901 GCAACACCCCAACATTAATAGGAGAGAAACAGAAACAGCTTATGACAAACAACTTGT 960  
| | | | |  
Db 876 GCAACACCCCAACATTAATAGGAGAGAAACAGAAACAGCTTATGACAAACAACTTGT 935  
| | | | |  
QY 961 GACCAACAGTGTATACAGATTTTGGAGGAGCAAGAGTTTGTCTGCTCTCACCGCT 1020  
| | | | |  
Db 936 GACCAACAGTGTATACAGATTTTGGAGGAGCAAGAGTTTGTCTGCTCTCACCGCT 995  
| | | | |  
QY 1021 GAGGGGATTAAGAACCCCAACAAACCTCAAGAGGCGCAGAGAGAGAGCGGCTTCCAAAT 1080  
| | | | |  
Db 996 GAGGGGATTAAGAACCCCAACAAACCTCAAGAGGCGCAGAGAGAGAGCGGCTTCCAAAT 1055  
| | | | |  
QY 1081 AACAGTAGCAGGCCCAAGACCCCAACATTAATGTGTCTGATCAAGAGATACAGACAGT 1140  
| | | | |  
Db 1056 AACAGTAGCAGGCCCAAGACCCCAACATTAATGTGTCTGATCAAGAGATACAGACAGT 1115  
| | | | |  
QY 1141 GATAGGAG 1200  
| | | | |  
Db 1116 GATAGGAG 1175  
| | | | |  
QY 1201 AAAAGTAGAACTTGAAGCTCTCTGAAGCAAAATTCGAGTCAAAACCAATTAAGATG 1260  
| | | | |  
Db 1176 AAAAGTAGAACTTGAAGCTCTCTGAAGCAAAATTCGAGTCAAAACCAATTAAGATG 1235  
| | | | |  
QY 1261 AAGCCAAATATTGAACCTCTCTGAGATGTGAGTGAAGTGTCTGAAGCCCTCAATGTTT 1320  
| | | | |  
Db 1236 AAGCCAAATATTGAACCTCTCTGAGATGTGAGTGAAGTGTCTGAAGCCCTCAATGTTT 1295  
| | | | |  
QY 1321 AGAGTCTCATTTGGCACTTAATGCAATTTCTGTGCCATTTGAGTTAATGGAGCC 1380  
| | | | |

Db 1296 AGAGTCTCATTTGGCACTTAATGCAATTTCTGTGCCATTTGAGTTAATGGAGCC 1355  
| | | | |  
QY 1381 AAAACATGATGACAGGTGTATGAGTTTGAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1440  
| | | | |  
Db 1356 AAAACATGATGACAGGTGTATGAGTTTGAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1415  
| | | | |  
QY 1441 CCGGCTGAGATGTGATTAATCTCTCCAGAGAAAGAGAGAGAAACCGGTTGTGGGCT 1500  
| | | | |  
Db 1416 CCGGCTGAGATGTGATTAATCTCTCCAGAGAAAGAGAGAGAAACCGGTTGTGGGCT 1475  
| | | | |  
QY 1501 GCAACATGCAAGAAAGATGACGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
| | | | |  
Db 1476 GCAACATGCAAGAAAGATGACGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535  
| | | | |  
QY 1536 CAACCTGTGATGATCAACCGGAGCTTGTGAGAGTTGAGTGTGATGATGACAA 1595  
| | | | |  
QY 1621 AATTTTGTGAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATG 1680  
| | | | |  
Db 1596 AATTTTGTGAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATG 1655  
| | | | |  
QY 1681 CGCTGCAAGAGACAGTGCACACCAAGCAGTGCCTGCTAAGCTGCTGCTGCTGCTGCTGCT 1740  
| | | | |  
Db 1656 CGCTGCAAGAGACAGTGCACACCAAGCAGTGCCTGCTAAGCTGCTGCTGCTGCTGCTGCT 1715  
| | | | |  
QY 1741 GACCTTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATTTGGACAGTAAATGTGTCC 1800  
| | | | |  
Db 1716 GACCTTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATTTGGACAGTAAATGTGTCC 1775  
| | | | |  
QY 1801 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGATCTATTTGCTGGACCACTTGAC 1860  
| | | | |  
Db 1776 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGATCTATTTGCTGGACCACTTGAC 1835  
| | | | |  
QY 1861 GTGGCAGGCTGGGGGATTTTATATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1920  
| | | | |  
Db 1836 GTGGCAGGCTGGGGGATTTTATATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1895  
| | | | |  
QY 1921 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTATGATTA 1980  
| | | | |  
Db 1896 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTATGATTA 1955  
| | | | |  
QY 1981 TACATGTGACGCTTCTGTTCAATGAAATTTTGTGTGTGATGCAACCCGCAAG 2040  
| | | | |  
Db 1956 TACATGTGACGCTTCTGTTCAATGAAATTTTGTGTGTGATGCAACCCGCAAG 2015  
| | | | |  
QY 2041 GGTAACAAAATTTGTTTGAAGATCATTCGGTAAATCAAACTGCTAATGCAAAAGTTATG 2100  
| | | | |  
Db 2016 GGTAACAAAATTTGTTTGAAGATCATTCGGTAAATCAAACTGCTAATGCAAAAGTTATG 2075  
| | | | |  
QY 2101 ATGTGTTAACGTTATGACAGATAGATATTTTTCGCAAGAGAGCCATCCAGACTGGCGAA 2160  
| | | | |  
Db 2076 ATGTGTTAACGTTATGACAGATAGATATTTTTCGCAAGAGAGCCATCCAGACTGGCGAA 2135  
| | | | |  
QY 2161 GAGCTGTTTATGATTAAGATTAAGCAAGCAAGGCTGATGCCCTGAAGTATGTGGGATGAA 2220  
| | | | |  
Db 2136 GAGCTGTTTATGATTAAGATTAAGCAAGCAAGGCTGATGCCCTGAAGTATGTGGGATGAA 2195  
| | | | |  
QY 2221 AGAGAAATGGAATCCCTTGA 2241  
| | | | |  
Db 2196 AGAGAAATGGAATCCCTTGA 2216  
| | | | |

Search completed: August 13, 2006, 13:10:52  
Job time : 2821 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2006, 11:58:10 ; Search time 522 Seconds  
(without alignments)  
6817.773 Million cell updates/sec

Title: US-10-773-302-1\_COPY\_90\_2330

Perfect score: 2241  
Sequence: 1 atgggcacgactgggaagaa.....gagaatggaatccctcga 2241

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2200221 seqs, 794037843 residues

Total number of hits satisfying chosen parameters: 4400442

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA New:\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_NEW\_PUB\_seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US06\_NEW\_PUB\_seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_NEW\_PUB\_seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_NEW\_PUB\_seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_NEW\_PUB\_seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_NEW\_PUB\_seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB\_seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB\_seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB\_seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US60\_NEW\_PUB\_seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2216	98.9	2655	US-11-266-748A-25769	Sequence 25769, A
2	2216	98.9	2695	US-10-539-630-3	Sequence 3, Appl1
3	2216	98.9	2695	US-11-266-748A-29501	Sequence 29501, A
4	2213	98.8	2253	US-10-539-630-2	Sequence 2, Appl1
5	1537.8	68.6	1803	US-11-266-748A-259286	Sequence 259286, A
6	1537.8	68.6	1803	US-11-266-748A-278870	Sequence 278870, A
7	1537.8	68.6	1803	US-11-266-748A-319803	Sequence 319803, A
8	980.8	43.8	1140	US-11-266-748A-98588	Sequence 98588, A
9	980.8	43.8	1140	US-11-266-748A-151399	Sequence 151399, A
10	940.6	42.0	2074	US-11-266-748A-186457	Sequence 186457, A
11	940.6	42.0	2074	US-11-266-748A-213375	Sequence 213375, A
12	840.4	37.5	4606	US-11-266-748A-62177	Sequence 62177, A
13	840.4	37.5	4606	US-11-266-748A-26094	Sequence 26094, A
14	837.2	37.4	1011	US-11-266-748A-175796	Sequence 175796, A
15	765.6	34.2	904	US-11-266-748A-361225	Sequence 361225, A
16	765.6	34.2	904	US-11-266-748A-444604	Sequence 444604, A
17	737	32.9	1000	US-11-266-748A-283007	Sequence 283007, A
18	737	32.9	1000	US-11-266-748A-30967	Sequence 30967, A
19	727	32.4	1000	US-11-266-748A-292953	Sequence 292953, A
20	727	32.4	1000	US-11-266-748A-344382	Sequence 344382, A
21	727	32.4	1000	US-11-266-748A-404820	Sequence 404820, A
22	727	32.4	1000	US-11-266-748A-475866	Sequence 475866, A
23	573	25.6	746	US-11-266-748A-357417	Sequence 357417, A

C 24	573	25.6	746	8	US-11-266-748A-440796	Sequence 440796, A
C 25	498	22.2	510	8	US-11-266-748A-8875	Sequence 8875, Ap
C 26	498	22.2	510	8	US-11-266-748A-63959	Sequence 63959, A
C 27	498	22.2	510	8	US-11-266-748A-66791	Sequence 66791, A
C 28	496.8	22.2	904	8	US-11-266-748A-216446	Sequence 216446, A
C 29	496.8	22.2	904	8	US-11-266-748A-238252	Sequence 238252, A
C 30	456.4	20.4	751	8	US-11-266-748A-74138	Sequence 74138, A
C 31	456.4	20.4	751	8	US-11-266-748A-108209	Sequence 108209, A
C 32	456.4	20.4	751	8	US-11-266-748A-126949	Sequence 126949, A
C 33	362	16.2	651	8	US-11-266-748A-355986	Sequence 355986, A
C 34	362	16.2	651	8	US-11-266-748A-439365	Sequence 439365, A
C 35	311.8	13.9	611	8	US-11-266-748A-263486	Sequence 263486, A
C 36	311.8	13.9	611	8	US-11-266-748A-324003	Sequence 324003, A
C 37	310.2	13.8	603	8	US-11-266-748A-14442	Sequence 14442, A
C 38	228	10.2	540	8	US-11-266-748A-42962	Sequence 42962, A
C 39	222	9.9	2313	8	US-11-266-748A-351035	Sequence 351035, A
C 40	222	9.9	2313	8	US-11-266-748A-383332	Sequence 383332, A
C 41	222	9.9	2313	8	US-11-266-748A-434414	Sequence 434414, A
C 42	188	8.4	897	8	US-11-266-748A-190098	Sequence 190098, A
C 43	179.2	8.0	449	8	US-11-266-748A-74139	Sequence 74139, A
C 44	179.2	8.0	449	8	US-11-266-748A-108210	Sequence 108210, A
C 45	179.2	8.0	449	8	US-11-266-748A-126950	Sequence 126950, A

## ALIGNMENTS

RESULT 1  
US-11-266-748A-25769  
Sequence 25769, Application US/11266748A  
GENERAL INFORMATION:  
APPLICANT: Harbin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcription Microarray Technology and  
FILE OF INVENTION: Methods of using the same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 48396  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 25769  
LENGTH: 2655  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-25769  
Query Match 98.9%; Score 2216; DB 8; Length 2655;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 15; Gaps 1;  
QY 1 ATGGCCGACTGGGAGAAATCTGAGAGGACCACTTTTGGCCGAGCGTATAAA 60  
DB 123 ATGGCCGACTGGGAGAAATCTGAGAGGACCACTTTTGGCCGAGCGTATAAA 182  
QY 61 TCAGAGTCAATGCGACTGAGACAGCTCAAGAGGTTTCAGACGACTGATGAAGTAAAGAT 120

Db	183	TCAGAGTACATGCGCATGAGCAGCTCAAGAGTTGAGACAGCATGTAAGTAAAGGT	242
Qy	121	ATGTTAGTTCCAAATCGTCAGAAAAATTTGGAAAAAGACGAATCTTTAAACCAAGATGG	180
Db	243	ATGTTAGTTCCAAATCGTCAGAAAAATTTGGAAAAAGACGAATCTTTAAACCAAGATGG	302
Qy	181	AAACAGCGAAGATACAGCTCTGTGCAATCTGCACTTCTGTAGCTCATTTGGCGGGAGCT	240
Db	303	AAACAGCGAAGATACAGCTCTGTGCAATCTGCACTTCTGTAGCTCATTTGGCGGGAGCT	362
Qy	241	AGGAGTGTGGGTGACCAAGTGAATTTTCCAAACAAGTCATCCATTAAAGACT	300
Db	363	AGGAGTGTGGGTGACCAAGTGAATTTTCCAAACAAGTCATCCATTAAAGACT	422
Qy	301	CTGAATGCAATGCTTCAGTACCACTAATATATTTCTGGTCTCCCTTACAGCAGAAATTTT	360
Db	423	CTGAATGCAATGCTTCAGTACCACTAATATATTTCTGGTCTCCCTTACAGCAGAAATTTT	482
Qy	361	ATGTGGAAGATGAACCTGTTTACATTAACATTCCTTATATATGGAGATGAAGTTTATGAT	420
Db	483	ATGTGGAAGATGAACCTGTTTACATTAACATTCCTTATATATGGAGATGAAGTTTATGAT	542
Qy	421	CAGATATGCTACTTTCATTTGAAGAATTAATAAAAAATTAATATATGGAAAGTACAGGGAGAT	480
Db	543	CAGATATGCTACTTTCATTTGAAGAATTAATAAAAAATTAATATATGGAAAGTACAGGGAGAT	602
Qy	481	AGAGATATGCGGTTATTAATATGATGAATTTTGTGAGATGATGAATGCCCTTGTCAA	540
Db	603	AGAGATATGCGGTTATTAATATGATGAATTTTGTGAGATGATGAATGCCCTTGTCAA	662
Qy	541	TATATATGATATGACGATGATGATGATGATGAGACGATCTGTAAGAAAGAAAGAAACAG	600
Db	663	TATATATGATATGACGATGATGATGATGATGAGACGATCTGTAAGAAAGAAAGAAACAG	722
Qy	601	AAAGATCTGGAGGATCACCGAGATGATTAAGAAAGCGGCCCACTCCGAAATTTCTCTTCT	660
Db	723	AAAGATCTGGAGGATCACCGAGATGATTAAGAAAGCGGCCCACTCCGAAATTTCTCTTCT	782
Qy	661	GATATAAATTTTGTGAAGCATTTTCTCATATGTTTCCAGATTAAGGCGACAGCAGAAAGACTA	720
Db	783	GATATAAATTTTGTGAAGCATTTTCTCATATGTTTCCAGATTAAGGCGACAGCAGAAAGACTA	842
Qy	721	AAGGAAAAATTAAGAACTCACCGGAACAGACGCTCCAGGCGCACTTCTCTTGATATGT	780
Db	843	AAGGAAAAATTAAGAACTCACCGGAACAGACGCTCCAGGCGCACTTCTCTTGATATGT	902
Qy	781	ACCCCGCAATATGATGACCAATATGCTAATATCTGTTCAGAGAGAGCAAAAGCTTACATCCC	840
Db	903	ACCCCGCAATATGATGACCAATATGCTAATATCTGTTCAGAGAGAGCAAAAGCTTACATCCC	962
Qy	841	TTTCAATAGCTTTTCTGTAGGCGATGTTTAAATATGATCTCTTCTCTCATC-----	892
Db	963	TTTCAATAGCTTTTCTGTAGGCGATGTTTAAATATGATCTCTTCTCTCATC-----	1022
Qy	893	-----CTTTCATGCAACACCCCAACTTATTAAGCGGAAGAACAGAAACAGCTCTA	945
Db	1023	AAATTAATCTTTCATGCAACACCCCAACTTATTAAGCGGAAGAACAGAAACAGCTCTA	1082
Qy	946	GACAAACAAACCTTGTGACACACATGTTTACAGCATTTTGGAGGAGCAAAAGATTTGCT	1005
Db	1083	GACAAACAAACCTTGTGACACACATGTTTACAGCATTTTGGAGGAGCAAAAGATTTGCT	1142
Qy	1006	GCTGTCTCACCGGTCGAGCGGATTAAGACCCCAACAAACGTTCCAGAGGCGCGAGAA	1065
Db	1143	GCTGTCTCACCGGTCGAGCGGATTAAGACCCCAACAAACGTTCCAGAGGCGCGAGAA	1202
Qy	1066	GGAGCGGCTTCCCAATTAACATGACAGGCGCCAGCACCCCAACATTAATATGTCTGGAATCA	1125
Db	1203	GGAGCGGCTTCCCAATTAACATGACAGGCGCCAGCACCCCAACATTAATATGTCTGGAATCA	1262
Qy	1126	AAAGTATCAGACATGATATGGAGACAGGACCTGAACCGGGGAGAGAACATATATAA	1185

Dp	1265	AAAGATATACAGACAGTGTATGAGGAAGCAGGACCTGAAAACGGGGGGAGAGAAACAATGATAAA	1322
Qy	1186	GAAGAGAAAGAGAGAAAGATGAAACTTCAGACTCCTCTGAAAGCAAAATTCCTGGTGTCAA	1245
Dp	1323	GAAGAAAGAAAGAGAAAGATGAAACTTCAGACTCCTCTGAAAGCAAAATTCCTGGTGTCAA	1382
Qy	1246	ACACCAATTAAGATGAAAGCCAAATATATGAAACCTCCCGAAGAAATGTGAGAGTGTGTCT	1305
Dp	1383	ACACCAATTAAGATGAAAGCCAAATATATGAAACCTCCCGAAGAAATGTGAGAGTGTGTCT	1442
Qy	1306	GAAGCCTCAATGTTTGAAGTCTCATATGGCACTTACCTATGACAAATTCCTGTGCCATTGCT	1365
Dp	1443	GAAGCCTCAATGTTTGAAGTCTCATATGGCACTTACCTATGACAAATTCCTGTGCCATTGCT	1502
Qy	1366	AGGTTAATTTGGAGCCAAAACATGTAGACAGGTGTATGAGTTTAAAGTCAAGAAATCTAGC	1425
Dp	1503	AGGTTAATTTGGAGCCAAAACATGTAGACAGGTGTATGAGTTTAAAGTCAAGAAATCTAGC	1562
Qy	1426	ATCATAGCTCCAGCTCCCGCTGAGATGTGAGTACCTCCCAAGGAAAAAAGAGAGAAA	1485
Dp	1563	ATCATAGCTCCAGCTCCCGCTGAGATGTGAGTACCTCCCAAGGAAAAAAGAGAGAAA	1622
Qy	1486	CACCGGTGTGGGCTGCACACTGECAGAAAAGATACAGCTGAAAAAGAGAGGCTCCTCTAAC	1545
Dp	1623	CACCGGTGTGGGCTGCACACTGECAGAAAAGATACAGCTGAAAAAGAGAGGCTCCTCTAAC	1682
Qy	1546	CATGTTTACAACTATCAACCTGTGTATCATCAACGAGACGCTTGTGACATTCGTGCTCT	1605
Dp	1683	CATGTTTACAACTATCAACCTGTGTATCATCAACGAGACGCTTGTGACATTCGTGCTCT	1742
Qy	1606	TGTGTGTATGACAAAAAATTTTGTGAAAAGTTTGTGCATGTATGTACAGAGTGTCAAAC	1665
Dp	1743	TGTGTGTATGACAAAAAATTTTGTGAAAAGTTTGTGTGCATGTATGTACAGAGTGTCAAAC	1802
Qy	1666	CGCTTTCCGGAGATGCGCTGCAGAAAGCAAGTGCACACCAAGCAGTGCCTGCTACCTG	1725
Dp	1803	CGCTTTCCGGAGATGCGCTGCAGAAAGCAAGTGCACACCAAGCAGTGCCTGCTACCTG	1862
Qy	1726	GCTGTCCGAGAGTGTGACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGAC	1785
Dp	1863	GCTGTCCGAGAGTGTGACCCCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGAC	1922
Qy	1786	AGTAAAAATGTGTCTCTGCGAAAGAACTGCACTATTCAGCCGGGGCTCCAAAAAGCATTTATG	1845
Dp	1923	AGTAAAAATGTGTCTCTGCGAAAGAACTGCACTATTCAGCCGGGGCTCCAAAAAGCATTTATG	1982
Qy	1846	CTGGCAACCATCTGACGTGCGAGGCTGTGGGGATTTTATACAAAGATCTGTGCGAGAAAAT	1905
Dp	1983	CTGGCAACCATCTGACGTGCGAGGCTGTGGGGATTTTATACAAAGATCTGTGCGAGAAAAT	2042
Qy	1906	GAAATTCATCTCAGAAATACTGTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGG	1965
Dp	2043	GAAATTCATCTCAGAAATACTGTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGG	2102
Qy	1966	AAAGTGTATATTAATAATACATGTGCAAGCTTTCGTTCACATTTGAACAAGATTTTGTGGT	2025
Dp	2103	AAAGTGTATATTAATAATACATGTGCAAGCTTTCGTTCACATTTGAACAAGATTTTGTGGT	2162
Qy	2026	GATGCAACCCGCAAGGGGTAAACAAAATTCGTTTTCGAAATCATTCGCTGTAATCTCAACTGC	2085
Dp	2163	GATGCAACCCGCAAGGGGTAAACAAAATTCGTTTTCGAAATCATTCGCTGTAATCTCAACTGC	2222
Qy	2086	TATGCAAAAGTTATATGATGTGTTAACGCTGTATCAAGAGTATGATATTTTTCGCAAGAGACC	2145
Dp	2223	TATGCAAAAGTTATATGATGTGTTAACGCTGTATCAAGAGTATGATATTTTTCGCAAGAGACC	2282
Qy	2146	ATTCAGACTGTGGCAGAGAGCTGTTTATGATACAGATACAGCCAGGCTGATGCCCTGAAG	2205
Dp	2283	ATTCAGACTGTGGCAGAGAGCTGTTTATGATACAGATACAGCCAGGCTGATGCCCTGAAG	2342
Qy	2206	TATGTGGGCATTCGAAAGAGAAATGGAATCCCTGTA	2241
Dp	2343	TATGTGGGCATTCGAAAGAGAAATGGAATCCCTGTA	2378

**RESULT 2**  
**INC-10-63**

```

US-10-539-630-3
: Sequence 3, Application US/10539630
: Publication No. US2006010499A1A1
: GENERAL INFORMATION:
: APPLICANT: Takeda Chemical Industries, Ltd.
: TITLE OF INVENTION: Preventing and treating agent for cancer
: FILE REFERENCE: 3130WOOP
: CURRENT APPLICATION NUMBER: US/10/539,630
: CURRENT FILING DATE: 2005-06-17
: PRIOR APPLICATION NUMBER: JP2002-373144
: PRIOR FILING DATE: 2002-12-24
: NUMBER OF SEQ ID NOS: 14
: SEQ ID NO 3
: LENGTH: 2695
: TYPE: DNA
: ORGANISM: Human
US-10-539-630-3

```

Query Match	98.9%	Score 2216;	DB 6;	Length 2695;
Best Local Similarity	99.3%	Pred. No. 0;		
Matches 2241; Conservative	0;	Mismatches 0;	Indels 15;	Gaps 1;

[illegible]

OY	721	AGGAAAAATTAAGAACTACCGAAGAGAGCTCCAGGCGCACTTCTCTAAAGT	780
DB	887	AAAGAAAAATTAAGAACTACCGAAGAGAGCTCCAGGCGCACTTCTCTAAAGT	946
OY	781	ACCCCAACATAGATGGAACCAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACCTCC	840
DB	947	ACCCCAACATAGATGGAACCAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACCTCC	1006
OY	841	TTTCAATAGCTTTTCTGTAGGCGAATGTTTTAAATATGATCTGTCTTCAATC-----	892
DB	1007	TTTCAATAGCTTTTCTGTAGGCGAATGTTTTAAATATGATCTGTCTTCAATC-----	1066
OY	893	-----CTTTCAAGCAACACCAACACTATATAGCGGAAGAACACAGAAACAGCTCTA	945
DB	1067	AATTAATCTTTTCAATGCAACCCCAACTATATAGCGGAAGAACACAGAAACAGCTCTA	1122
OY	946	GACAAACAACCTTGTGACCAAGTGTATCCAGCAATTTGAGGAGCAAAAGAGTTTGTCT	1005
DB	1127	GACAAACAACCTTGTGACCAAGTGTATCCAGCAATTTGAGGAGCAAAAGAGTTTGTCT	1188
OY	1006	GCTGTCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTCCAGAGGCGCGAGAA	1065
DB	1187	GCTGTCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTCCAGAGGCGCGAGAA	1246
OY	1066	GGAAGGCTTCCCAATTAACATGAGAGGCGCCAGACCCCAACATTAATATGTCTGGAATCA	1122
DB	1247	GGAAGGCTTCCCAATTAACATGAGAGGCGCCAGACCCCAACATTAATATGTCTGGAATCA	1306
OY	1126	AAGATATCAGACAGTGTATAGGGAAGCAGAGCACTGAAACGGGGGAGAGAACATATGATAA	1185
DB	1307	AAGATATCAGACAGTGTATAGGGAAGCAGAGCACTGAAACGGGGGAGAGAACATATGATAA	1366
OY	1186	GAAGAGAAGAGAAAGATGAACTTCAGACTCTCTGAAGCAATTCCTCGGTGCAA	1245
DB	1367	GAAGAGAAGAGAGAAAGATGAACTTCAGACTCTCTGAAGCAATTCCTCGGTGCAA	1428
OY	1246	ACACCAATTAAGATGAAGCCCAATATTTGAACCTCTCTGAAATGTGTAGATGAGATGTCT	1305
DB	1427	ACACCAATTAAGATGAAGCCCAATATTTGAACCTCTCTGAAATGTGTAGATGAGATGTCT	1486
OY	1306	GAAGCTCAATGTTTAAAGTCTCTCAATTGGGCACTTATGACAAATTCCTGTGCGCATTTGCT	1365
DB	1487	GAAGCTCAATGTTTAAAGTCTCTCAATTGGGCACTTATGACAAATTCCTGTGCGCATTTGCT	1544
OY	1366	AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAAAGTCAAAAGATCTAGC	1425
DB	1547	AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAAAGTCAAAAGATCTAGC	1606
OY	1426	ATCATAGTCTCAGTCCCGCTGAGAGATGTGATCTCTCCAGAGAAAAGAGAGGAAA	1485
DB	1607	ATCATAGTCTCAGTCCCGCTGAGAGATGTGATCTCTCCAGAGAAAAGAGAGGAAA	1666
OY	1486	CACCGGTTTGGGGCTGACACTGAGAAAGTACAGCTGAAAAAAGAGACGGCTCCCTTAC	1545
DB	1667	CACCGGTTTGGGGCTGACACTGAGAAAGTACAGCTGAAAAAAGAGACGGCTCCCTTAC	1726
OY	1546	CATGTTTAACTATCAACCTGTGATCATCAAGGAGCCTTGTGAACAGTTCTGTGCGCT	1605
DB	1727	CATGTTTAACTATCAACCTGTGATCATCAAGGAGCCTTGTGAACAGTTCTGTGCGCT	1786
OY	1606	TGTGTATGACAAAAATTTTGTGAAAAGTTTGTCAATGTACTTCAAGATGTCAAAAC	1665
DB	1787	TGTGTATGACAAAAATTTTGTGAAAAGTTTGTCAATGTACTTCAAGATGTCAAAAC	1846
OY	1666	CGCTTTTCCGGAGTCCGCTGTCAAAAGCAAGTGAACAACAAGTGCCTGTGTAACCTG	1725
DB	1847	CGCTTTTCCGGAGTCCGCTGTCAAAAGCAAGTGAACAACAAGTGCCTGTGTAACCTG	1906
OY	1726	GCTGTCCGAGAGTGTACCTGTGACCTCTGTCTTACTGTGTGAGGCGCGTGAACATTTGGGAC	1785
DB	1907	GCTGTCCGAGAGTGTACCTGTGACCTCTGTCTTACTGTGTGAGGCGCGTGAACATTTGGGAC	1966
OY	1786	AGTAAAAATGTCTCTGCAAGAACTGCAATTCAGCGGGGCTCAAAAAGCATCTATTG	1845

Db 1967 AGTAAATGTCTCTGCAAGAACTGACAGTATTGACGGGGCTCCAAAAGCATCTATTG 2026  
Qy 1646 CTGGCAACATCTGACGTGGCAAGGCTGGGGGATTTTATCAAGATCCGTGACAGAAAT 1905  
Db 2027 CTGGCAACATCTGACGTGGCAAGGCTGGGGGATTTTATCAAGATCCGTGACAGAAAT 2086  
Qy 1906 GAATTCATCTGAGATATCTGTGAGAGATTTTCTCAAGATGAGTGAAGAGAGG 1965  
Db 2087 GAATTCATCTGAGATATCTGTGAGAGATTTTCTCAAGATGAGTGAAGAGAGG 2146  
Qy 1966 AAGGTATGATTAATATACATGTGACGCTTTCTGTTCAATCTGAACATGATTTGTGTG 2025  
Db 2147 AAGGTATGATTAATATACATGTGACGCTTTCTGTTCAATCTGAACATGATTTGTGTG 2206  
Qy 2026 GATGCAACCCCAAGGTTAAACAAATTTGTTTGAATTCATTTGGTAAATCCAAAGCC 2085  
Db 2207 GATGCAACCCCAAGGTTAAACAAATTTGTTTGAATTCATTTGGTAAATCCAAAGCC 2266  
Qy 2086 TATGCAAAAGTTATGATGTTTAAACGATGATCAAGATAGGTATTTTCCAAAGAGCC 2145  
Db 2267 TATGCAAAAGTTATGATGTTTAAACGATGATCAAGATAGGTATTTTCCAAAGAGCC 2326  
Qy 2146 ATCCAGACTGGCAAGAGCTGTTTGTATTCAGATACAGCCAGGCTGATGCTTGAAG 2205  
Db 2327 ATCCAGACTGGCAAGAGCTGTTTGTATTCAGATACAGCCAGGCTGATGCTTGAAG 2386  
Qy 2206 TATGCGGATCGAAAGAAATGAGAAATCCCTTGA 2241  
Db 2387 TATGCGGATCGAAAGAAATGAGAAATCCCTTGA 2422

## RESULT 3

US-11-266-748A-29501  
Sequence 29501, Application US/11266748A  
Publication No. US2006034663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcription Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 29501  
LENGTH: 2695  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-29501

Query Match 98.9%; Score 2216; DB 8; Length 2695;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2241; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

Qy 1 ATGGCCAGACCTGGGAAGAAATCTGAGAGGACCAATTTGTCGGGAAGCCGTGTA 60  
Db 167 ATGGCCAGACCTGGGAAGAAATCTGAGAGGACCAATTTGTCGGGAAGCCGTGTA 226  
Qy 61 TCAGAGTACATGCGACCTGAGACAGCTCAAGGGTTGAGAGAGTGAAGTAAAGT 120  
Db 227 TCAGAGTACATGCGACCTGAGACAGCTCAAGGGTTGAGAGAGTGAAGTAAAGT 286  
Qy 121 ATGTTAGTTCATCTGAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 180  
Db 287 ATGTTAGTTCATCTGAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 346  
Qy 181 AAACGAGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGGGAGT 240  
Db 347 AAACGAGAGAGATACAGCTGTGACATCTGACTTCTGTGAGCTCATTTGGGGAGT 406  
Qy 241 AGGAGGTTCGAGACCAAGTCACTTGAATTTTCAACAGATCATCCATTAAAGT 300  
Db 407 AGGAGGTTCGAGACCAAGTCACTTGAATTTTCAACAGATCATCCATTAAAGT 466  
Qy 301 CTGATGCAAGTGTCTGATACCAATATGATTTCTGTCTCCCTTACAGAGATTTT 360  
Db 467 CTGATGCAAGTGTCTGATACCAATATGATTTCTGTCTCCCTTACAGAGATTTT 526  
Qy 361 ATGTGGAAGATGAACCTGTTTATCATTAATCTTATATGGAAGATGAAGTTTATG 420  
Db 527 ATGTGGAAGATGAACCTGTTTATCATTAATCTTATATGGAAGATGAAGTTTATG 586  
Qy 421 CAGAGTGTACTTTCATTGAAAGAACTAATAAATATGATGGAAGATGACAGGGAT 480  
Db 587 CAGAGTGTACTTTCATTGAAAGAACTAATAAATATGATGGAAGATGACAGGGAT 646  
Qy 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGATGCCCTTGTCTAA 540  
Db 647 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTGTGATGCCCTTGTCTAA 706  
Qy 541 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Db 707 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
Qy 601 AAAGATGTGGAGATCACCGAGATGATTAAGAAAGCCGCCAATCTCGAAATTTCTTCT 660  
Db 767 AAAGATGTGGAGATCACCGAGATGATTAAGAAAGCCGCCAATCTCGAAATTTCTTCT 826  
Qy 661 GATTAATTTTGAAGCATTTCTCAATGTTTCAATTAAGGAGACAGAGAACTA 720  
Db 827 GATTAATTTTGAAGCATTTCTCAATGTTTCAATTAAGGAGACAGAGAACTA 886  
Qy 721 AAGGAAATTTTAAGAACTACCGAACAGAGCTCCCAAGGCGAATTTCTCGAATGT 780  
Db 887 AAGGAAATTTTAAGAACTACCGAACAGAGCTCCCAAGGCGAATTTCTCGAATGT 946  
Qy 781 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTGAGAGAGCAAAAGCTTCACTCC 840  
Db 947 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTGAGAGAGCAAAAGCTTCACTCC 1006  
Qy 841 TTTCATAGCTTTTCTGAGGAGATGTTTAAATATGATGCTTCTTCAATC----- 892  
Db 1007 TTTCATAGCTTTTCTGAGGAGATGTTTAAATATGATGCTTCTTCAATC----- 1066  
Qy 893 -----CTTTTATGCAACCCCAACCTATTAAGGGGAAGAACAGAAAGCTCTA 945  
Db 1067 AATTATCTTTTATGCAACCCCAACCTATTAAGGGGAAGAACAGAAAGCTCTA 1126  
Qy 946 GACAACAAACCTTGTGACACAGTGTATCAAGATTTGAGAGAGCAAAAGCTTTGCT 1005  
Db 1127 GACAACAAACCTTGTGACACAGTGTATCAAGATTTGAGAGAGCAAAAGCTTTGCT 1186  
Qy 1006 GCTGCTCTCAACGCTGAGCGGATTAAGACCCCAACAAACCTTCAAGAGGCGCAAGAA 1065  
Db 1187 GCTGCTCTCAACGCTGAGCGGATTAAGACCCCAACAAACCTTCAAGAGGCGCAAGAA 1246  
Qy 1066 GAGCGGCTTCCCAATTAACAGTACAGGCGCCAGCACCCCAACATTATGTGCTGAATCA 1125

Db	1247	GGACGGCTTCCCAATAAACATGACGAGCCCAACACCCCACTTAATGTCTGGAAATCA	1306
Oy	1126	AAGATATACAGACAGTGAATAGGGAGACAGGACCTGAACCGGGGGAGAGAACATATGATAA	1185
Db	1307	AAGGATATACAGACAGTATAGGAGACAGGACCTGAACCGGGGGAGAGAACATATGATAA	1366
Oy	1186	GAAGAGAGAGAGAGAGAAATGATAAATTCTGAGCTCTCTGAAGCAAAATTCCTGGTGTCA	1245
Db	1367	GAAGAGAGAGAGAGAGAAATGATAAATTCTGAGCTCTCTGAAGCAAAATTCCTGGTGTCA	1426
Oy	1246	ACACCAATTAAGAATGAGCCAAATATTAACCTCTCGAAGATGTGAGATGAGTGCT	1305
Db	1427	ACACCAATTAAGAATGAGCCAAATATTAACCTCTCGAAGATGTGAGATGAGTGAGTCT	1486
Oy	1306	GAAGCTCAATGTTTAGATGCTCTCAATTTGCACTTAATGACAAATTCGTGCCATTGCT	1365
Db	1487	GAAGCTCAATGTTTAGATGCTCTCAATTTGCACTTAATGACAAATTCGTGCCATTGCT	1546
Oy	1366	AGGTTAATTTGGACCAAAACATGTAGACAGGTATGAGTTTAGAGTCAAAAGATCTAGC	1425
Db	1547	AGGTTAATTTGGACCAAAACATGTAGACAGGTATGAGTTTAGAGTCAAAAGATCTAGC	1606
Oy	1426	ATCATAGCTCAGGCTCCGCTGAGGAGATGTGATCTCCTCCAGGAAAATAAGAGGAA	1485
Db	1607	ATCATAGCTCAGGCTCCGCTGAGGAGATGTGATCTCCTCCAGGAAAATAAGAGGAA	1666
Oy	1486	CACCGGTGTGGGCTGCACACTGCGAGAAAGATACAGCTGAAAAAGACCGCTCTCTAAC	1545
Db	1667	CACCGGTGTGGGCTGCACACTGCGAGAAAGATACAGCTGAAAAAGACCGCTCTCTAAC	1726
Oy	1546	CATGTTTACAATATCAACCTGTGATCATCAACGCGACGCTTGTGACAGTTCGGTCCCT	1605
Db	1727	CATGTTTACAATATCAACCTGTGATCATCAACGCGACGCTTGTGACAGTTCGGTCCCT	1786
Oy	1606	TGTGTGATAGACAAAAATTTTGTGAAAAGTTTGTCAATGTAGTTCAGATGTCAAAAC	1665
Db	1787	TGTGTGATAGACAAAAATTTTGTGAAAAGTTTGTCAATGTAGTTCAGATGTCAAAAC	1846
Oy	1666	CGCTTTCCGGAGATGCGCTGCAAAAGCACAGTGCACACCAAGCAGTGCCTGTGTAACCTG	1725
Db	1847	CGCTTTCCGGAGATGCGCTGCAAAAGCACAGTGCACACCAAGCAGTGCCTGTGTAACCTG	1906
Oy	1726	GCTGTCCGAGAGTGTGACCTGTGCTTACTGTGTGAGACCGCTGACATTTGGAC	1785
Db	1907	GCTGTCCGAGAGTGTGACCTGTGCTTACTGTGTGAGACCGCTGACATTTGGAC	1966
Oy	1786	AGTAAAAATGTGTCTGTGCMAAACCTGCGATTTCAAGCGGGCTCAAAAAGCATCTATTG	2026
Db	1967	AGTAAAAATGTGTCTGTGCMAAACCTGCGATTTCAAGCGGGCTCAAAAAGCATCTATTG	
Oy	1846	CTGGCACATCTGACGTGGCAGGCGCTGGGGGATTTTATCAAGATTCGTGCAAAAAAT	1905
Db	2027	CTGGCACATCTGACGTGGCAGGCGCTGGGGGATTTTATCAAGATTCGTGCAAAAAAT	2086
Oy	1906	GAATTCATCTCAGATACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGG	1965
Db	2087	GAATTCATCTCAGATACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGG	2146
Oy	1966	AAAGTGTATGATAAATATCATGTGACGCTTCTGTTCAACTTGAACAATGATTTGTGATG	2025
Db	2147	AAAGTGTATGATAAATATCATGTGACGCTTCTGTTCAACTTGAACAATGATTTGTGATG	2206
Oy	2026	GATGCAACCCGCAAGGGTAAACAAAATTCGTTTGCAATCATTTGGGTAATCCAAACCTGC	2085
Db	2207	GATGCAACCCGCAAGGGTAAACAAAATTCGTTTGCAATCATTTGGGTAATCCAAACCTGC	2266
Oy	2086	TATGCAAAAGTTATGATGTTTAAACGATATCACAGATAGGTTATTTTGTCCAAAGAGCC	2145
Db	2267	TATGCAAAAGTTATGATGTTTAAACGATATCACAGATAGGTTATTTTGTCCAAAGAGCC	2326
Oy	2146	ATTCAGACTGCGAGAGGCTGTTTGTATTAACAGATACAGCCAGGCTGATGCCCTGAG	2205

[illegible]

QY 661 GATATAATTTTGAAGCATTCTCAATGTTTCCAGATAAGGGCAACAGAAAGAACTA 720  
| | | | |  
Db 661 GATATAATTTTGAAGCATTCTCAATGTTTCCAGATAAGGGCAACAGAAAGAACTA 720  
QY 721 AAGGAAAAATTAAGAACTCAACGAAACAGAGCTCCAGGGGCACTTCCCTCGAATGT 780  
| | | | |  
Db 721 AAGGAAAAATTAAGAACTCAACGAAACAGAGCTCCAGGGGCACTTCCCTCGAATGT 780  
QY 781 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACCTCC 840  
| | | | |  
Db 781 ACCCCCAATAGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACCTCC 840  
QY 841 TTTCAATGCTTTTCTGAGCGATGTTTAAATATATGATCTGCTTCAATC----- 892  
| | | | |  
Db 841 TTTCAATGCTTTTCTGAGCGATGTTTAAATATATGATCTGCTTCAATC----- 892  
QY 893 -----CTTTTCAATGCAACCCCAACTTNTAAGCGGAGAAACACAGAAACAGCTCTA 945  
| | | | |  
Db 901 AATTATCTTTTCAATGCAACCCCAACTTNTAAGCGGAGAAACACAGAAACAGCTCTA 960  
QY 946 GACAACAAACCTTGTGACACAGTGTACAGCAATTTGAGAGAGCAAAAGATTGCT 1005  
| | | | |  
Db 961 GACAACAAACCTTGTGACACAGTGTACAGCAATTTGAGAGAGCAAAAGATTGCT 1020  
QY 1006 GCTGCTCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTTCAAGAGCGCGAAGAA 1065  
| | | | |  
Db 1021 GCTGCTCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTTCAAGAGCGCGAAGAA 1080  
QY 1066 GGAAGCGCTTCCCAATACAGTAAGAGCGGCGACCCCACTTAATATGTCGTAATCA 1125  
| | | | |  
Db 1081 GGAAGCGCTTCCCAATACAGTAAGAGCGGCGACCCCACTTAATATGTCGTAATCA 1140  
QY 1126 AAGGATACAGACAGTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1185  
| | | | |  
Db 1141 AAGGATACAGACAGTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1186 GAAG 1245  
| | | | |  
Db 1201 GAAG 1260  
QY 1246 ACACCAATTAAGATGAAGCCAAATTTGAACCTCTGAGAGATGAGATGAGATGAGATG 1305  
| | | | |  
Db 1261 ACACCAATTAAGATGAAGCCAAATTTGAACCTCTGAGAGATGAGATGAGATGAGATG 1320  
QY 1306 GAAGCCTCAATGTTTGAAGCTCTCAATGAGCACTTACTATGACAAATTTCTGTCATGCT 1365  
| | | | |  
Db 1321 GAAGCCTCAATGTTTGAAGCTCTCAATGAGCACTTACTATGACAAATTTCTGTCATGCT 1380  
QY 1366 AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGC 1425  
| | | | |  
Db 1381 AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGC 1440  
QY 1426 ATCAATAGCTCAGCTCCGCTGAGAGATGATGATCTCTCCAGAGAGAAAGAGAGAGAA 1485  
| | | | |  
Db 1441 ATCAATAGCTCAGCTCCGCTGAGAGATGATGATCTCTCCAGAGAGAAAGAGAGAGAA 1500  
QY 1486 CACCGGTTGTGGGCTGACACTGACAGAAAGATACAGCTGAGAAAGAGAGAGAGAGAGAG 1545  
| | | | |  
Db 1501 CACCGGTTGTGGGCTGACACTGACAGAAAGATACAGCTGAGAAAGAGAGAGAGAGAG 1560  
QY 1546 CATGTTTACAACTATCAACCTGTGATCATCAACGAGAGCTTGTGACAGTTGCTGCTT 1605  
| | | | |  
Db 1561 CATGTTTACAACTATCAACCTGTGATCATCAACGAGAGCTTGTGACAGTTGCTGCTT 1620  
QY 1606 TGTGTGAATGACAAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAAC 1665  
| | | | |  
Db 1621 TGTGTGAATGACAAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAAC 1680  
QY 1666 CGCTTTCCGGGATGCGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1725  
| | | | |  
Db 1681 CGCTTTCCGGGATGCGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

QY 1726 GCTGTCCGAGAGATGTGACCTGTGACCTTGTGCTTACTTGTGAGAGCCGCTGACCAATGGGAC 1785  
| | | | |  
Db 1741 GCTGTCCGAGAGATGTGACCTGTGACCTTGTGCTTACTTGTGAGAGCCGCTGACCAATGGGAC 1800  
QY 1786 AGTAAAAATGTGTCCTGACAGAAAGTGTGAGATTCAGCGGGGCTCCAAAAAGCATCTATTG 1845  
| | | | |  
Db 1801 AGTAAAAATGTGTCCTGACAGAAAGTGTGAGATTCAGCGGGGCTCCAAAAAGCATCTATTG 1860  
QY 1846 CTGGACCATCTGACGCTGAGCGAGCTGGGGATTTTATCAAGATCTGTGCAAGAAAT 1905  
| | | | |  
Db 1861 CTGGACCATCTGACGCTGAGCGAGCTGGGGATTTTATCAAGATCTGTGCAAGAAAT 1920  
QY 1906 GAATTCATCTCAAAATCTGTGAGAGATTAATTTTCAAGATGAGAGTGAACAGAGAGG 1965  
| | | | |  
Db 1921 GAATTCATCTCAAAATCTGTGAGAGATTAATTTTCAAGATGAGAGTGAACAGAGAGG 1980  
QY 1966 AAGGTGATGATTAATTCATGTGCAAGCTTCTGTTCAACTTGAACATGATTTTGTGCTG 2025  
| | | | |  
Db 1981 AAGGTGATGATTAATTCATGTGCAAGCTTCTGTTCAACTTGAACATGATTTTGTGCTG 2040  
QY 2026 GATGCAACCCGCAAGGGTAAACAAATTCGTTTGCAAATCATTCGTAATCCAACTGC 2085  
| | | | |  
Db 2041 GATGCAACCCGCAAGGGTAAACAAATTCGTTTGCAAATCATTCGTAATCCAACTGC 2100  
QY 2086 TATGCAAAATGTATGATGCTTTAAGGTATCAAGATCAAGATAGATATTTTGCAGAGAGCC 2145  
| | | | |  
Db 2101 TATGCAAAATGTATGATGCTTTAAGGTATCAAGATCAAGATAGATATTTTGCAGAGAGCC 2160  
QY 2146 ATCCAGACTGGCGAAGAGCTGTTTTTGAATTAAGATACAGACAGGCTGATGCTGTAAG 2205  
| | | | |  
Db 2161 ATCCAGACTGGCGAAGAGCTGTTTTTGAATTAAGATACAGACAGGCTGATGCTGTAAG 2220  
QY 2206 TATGTGGCATCGAAAGAGAAATGAAATCCCT 2238  
| | | | |  
Db 2221 TATGTGGCATCGAAAGAGAAATGAAATCCCT 2253  
| | | | |

RESULT 5  
US-11-266-748A-259286  
; Sequence 259286, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Hartlin, Paul  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT FILING DATE: US/11/266, 748A  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662, 276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700, 293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 48396  
; SOFTWARE: Patent version 3.3  
; SEQ ID NO 259286  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:



NAME/KEY: misc feature  
LOCATION: (356) .. (379)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1680) .. (1680)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1764) .. (1764)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1792) .. (1792)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-259286

Query Match 68.6%; Score 1537.8; DB 8; Length 1803;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 1616; Conservative 0; Mismatches 26; Indels 7; Gaps 7;

QY 269 ATTTTCCACACAGTCATCCCATTAAGACTCTGAATGCAAGTGGCTTCAAGTACCATA 328  
DB 80 ATTTCCGACACAGTCATCCCATTAAGACTCTGAATGCAAGTGGCTTCAAGTACCATA 139  
QY 329 TGTATTTCTTGTCTCCCTACAGCAAGATTTATGTGGAAGATGAATCTGTTTACATA 388  
DB 140 TGTATTTCTTGTCTCCCTACAGCAAGATTTATGTGGAAGATGAATCTGTTTACATA 199  
QY 389 ACATTCCTTAATATGGAAGATGAAGTGTATGATCAGAGTGTACTTTCATTGAAGACTTA 448  
DB 200 ACATTCCTTAATATGGAAGATGAAGTGTATGATCAGAGTGTACTTTCATTGAAGACTTA 259  
QY 449 TAAAAAATTATGATGGAAGATGAAGTGTATGATCAGAGTGTACTTTCATTGAAGACTTA 508  
DB 260 TAAAAAATTATGATGGAAGATGAAGTGTATGATCAGAGTGTACTTTCATTGAAGACTTA 319  
QY 509 TTTTGTGTGAGTGTGATGATGAGTGTGATGATGATGATGATGATGATGATGATGATG 568  
DB 320 TTTTGTGTGAGTGTGATGATGAGTGTGATGATGATGATGATGATGATGATGATGATG 379  
QY 569 GAGACGATCTCTGAAGAGAGAGAGAGAGAGAGATCTGAGAGATCAACGAGATGATA 628  
DB 380 GAGACGATCTCTGAAGAGAGAGAGAGAGAGAGATCTGAGAGATCAACGAGATGATA 439  
QY 629 AAGAAAGCGCCGACCTCGGAAATTTCTTCTGATTAATTTTGAAGCATTTCTCAA 688  
DB 440 AAGAAAGCGCCGACCTCGGAAATTTCTTCTGATTAATTTTGAAGCATTTCTCAA 499  
QY 689 TGTTCAGATTAAGGAGAGAGAGAGAGAGAGATTAAGAGAGAGATTAAGAGAGAGAG 748  
DB 500 TGTTCAGATTAAGGAGAGAGAGAGAGAGAGATTAAGAGAGAGATTAAGAGAGAGAG 559  
QY 749 AGCAGCTCCAG 808  
DB 560 AGCAGCTCCAG 619  
QY 809 AATCTGTTCAG 868  
DB 620 AATCTGTTCAG 679  
QY 869 TTAATATGAG 928  
DB 680 TTAATATGAG 739  
QY 929 ACACAGAAACAGCTCTAGACAAACCTTGTGACACAGAGTATACAGAGATTTGAGAG 988  
DB 740 ACACAGAAACAGCTCTAGACAAACCTTGTGACACAGAGTATACAGAGATTTGAGAG 799  
QY 989 GAGCAAG 1048  
DB 800 GAGCAAG 859

QY 1049 CAGAGAGCGGAG 1108  
DB 860 CAGAGAGCGGAG 919  
QY 1109 TTAATGTCTGGAATCAAGAGATCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 1168  
DB 920 TTAATGTCTGGAATCAAGAGATCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 979  
QY 1169 GAGAGAAATGATTAAG 1228  
DB 980 GAGAGAAATGATTAAG 1039  
QY 1229 CAAATTCCTGAGTCAAAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1288  
DB 1040 CAAATTCCTGAGTCAAAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1099  
QY 1289 TGAAGTGAAGTGTGCTGAAGCTCAATGTTTGAAGTCTTCACTTGAAGTGAAG 1348  
DB 1100 TGAAGTGAAGTGTGCTGAAGCTCAATGTTTGAAGTCTTCACTTGAAGTGAAG 1159  
QY 1349 ATTTCTGTGCAATGCTGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1408  
DB 1160 ATTTCTGTGCAATGCTGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219  
QY 1409 GAGTCAAGAAATCTAGATCAATGCTGAGTCCGCTGAGAGATGATTAATCTCCAA 1468  
DB 1220 GAGTCAAGAAATCTAGATCAATGCTGAGTCCGCTGAGAGATGATTAATCTCCAA 1279  
QY 1469 GAGAAAG 1527  
DB 1280 GAGAAAG 1339  
QY 1528 AAGAGCGCTCTCTTAACAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1586  
DB 1340 AAGAGCGCTCTCTTAACAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1399  
QY 1587 TGTGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1645  
DB 1400 TGTGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1459  
QY 1646 GTAGTT-CAGAGTCAAAACCGCTTT-CGAGAGTCCGCTGCAAGAGAGAGAGAGAG 1703  
DB 1460 GTAGTTCAAGAGTCAAAACCGCTTTCCGAGAGTCCGCTGCAAGAGAGAGAGAGAG 1519  
QY 1704 CAGAGAGTCCGCTGCTA-CCTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1762  
DB 1520 CAGAGAGTCCGCTGCTA-CCTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1579  
QY 1763 GTGAGCGCTGAG 1822  
DB 1580 GTGAGCGCTGAG 1639  
QY 1823 GGGGCTCCAAAAGAGATTAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881  
DB 1640 GGGGCTCCAAAAGAGATTAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699  
QY 1882 ATCAAGATCTGTGAG 1910  
DB 1700 ATCAAGATCTGTGAG 1728

RESULT 6  
US-11-266-748A-278870  
; Sequence 278870, Application US/11266748A  
; Publication No. US2006013463A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnston, Paul  
; APPLICANT: Mulligan, Kari  
; TITLE OF INVENTION: Transcription Microarray Technology and  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A



Db 1580 GTGAGCCGCTGACCATTTGGGACAGTAAATGTGCTCGCAAGAACTGATTTGACG 1639  
Qy 1823 GGGGCTCCAAAAGCATTTATTTGCTGGCACTTCGACGT-GGCAAGCTGGGGGATTTTT 1881  
Db 1640 GGGGCTCCAAAAGCATTTATTTGCTGGCACTTCGACAGCTGGGGGATTTTT 1699  
Qy 1882 ATCAAGATCTGTGAGAAAATGAATT 1910  
Db 1700 ATCAAGATCTGTGAGAAAATGAATT 1728

RESULT 7  
US-11-266-748A-319803/C  
; Sequence 319803, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIOR FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 48396  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 319803  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (12)..(12)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40)..(40)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (124)..(124)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1425)..(1448)  
; OTHER INFORMATION: n is a, c, g, or t  
; US-11-266-748A-319803

Query Match 68.6%; Score 1537.8; DB 8; Length 1803;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 1616; Conservative 0; Mismatches 26; Indels 7; Gaps 7;

Qy 269 ATTTCCACACAGTCATCCCATTAAGACTCTGATGACAGTTGCTTCAGTACCATTA 328  
Db 1724 ATTCCGACACAGTCATCCCATTAAGACTCTGATGACAGTTGCTTCAGTACCATTA 1665

Qy 329 TGTATCTTGGTCTCCCTCAGACAGAAATTTATGATGGAAGATGAACCTGTTTACATA 388  
Db 1664 TGTATCTTGGTCTCCCTCAGACAGAAATTTATGATGGAAGATGAACCTGTTTACATA 1605  
Qy 389 ACATTCCTTAATATGGAAGATGAAGTTTATGATCAGATGATGATCTTCAATGGAAGATTA 448  
Db 1604 ACATTCCTTAATATGGAAGATGAAGTTTATGATCAGATGATGATCTTCAATGGAAGATTA 1545  
Qy 449 TAAAAATTTATGATGGAAGATGAACGGGGATGGAATGTGGGTTTATTAATGATGA 508  
Db 1544 TAAAAATTTATGATGGAAGATGAACGGGGATGGAATGTGGGTTTATTAATGATGA 1485  
Qy 509 TTTTGTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 568  
Db 1484 TTTTGTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1425  
Qy 569 GAGACGATCTGGAAGAAAGAAAGAAAGAAAGATCTGAGAGATCAACGAGATGATA 628  
Db 1424 GAGACGATCTGGAAGAAAGAAAGAAAGAAAGATCTGAGAGATCAACGAGATGATA 1365  
Qy 629 AAGAAAGCCGCCACCTCGGAATTTCTCTGATTAATTTTGAAGCATTTCTCTCA 688  
Db 1364 AAGAAAGCCGCCACCTCGGAATTTCTCTGATTAATTTTGAAGCATTTCTCTCA 1305  
Qy 689 TGTTCAGATTAAGGCAAGAGAAAGAAAGAAAGAAATTAAGAACTCACGAAAC 748  
Db 1304 TGTTCAGATTAAGGCAAGAGAAAGAAAGAAAGAAATTAAGAACTCACGAAAC 1245  
Qy 749 AGACGCTCCAGGCGCACTTCTCTGATGATGATGATGATGATGATGATGATGATGAT 808  
Db 1244 AGACGCTCCAGGCGCACTTCTCTGATGATGATGATGATGATGATGATGATGATGAT 1185  
Qy 809 AATCTGTTGAGAGAGAAAGCTTAACCTCTTCAATGATGATGATGATGATGATGATG 868  
Db 1184 AATCTGTTGAGAGAGAAAGCTTAACCTCTTCAATGATGATGATGATGATGATGATG 1125  
Qy 869 TTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 928  
Db 1124 TTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1065  
Qy 929 ACACAGAAACAGCTCTAGACAAACCTTGTGACCAAGTGTACAGACATTTGAGAG 988  
Db 1064 ACACAGAAACAGCTCTAGACAAACCTTGTGACCAAGTGTACAGACATTTGAGAG 1005  
Qy 989 GAGCAAGAGATTTGCTGCTGCTCTCAACGCTGAGCGGATTAAGACCCCAACCAACGTC 1048  
Db 1004 GAGCAAGAGATTTGCTGCTGCTCTCAACGCTGAGCGGATTAAGACCCCAACCAACGTC 945  
Qy 1049 CAGAGGCGCGAAG 1108  
Db 944 CAGAGGCGCGAAG 885  
Qy 1109 TTAATGCTGGAATCAAGAGATACAGACAGTATGAGAGAGAGAGAGAGAGAGAGAG 1168  
Db 884 TTAATGCTGGAATCAAGAGATACAGACAGTATGAGAGAGAGAGAGAGAGAGAGAG 825  
Qy 1169 GAGAGAACATGATTAAG 1228  
Db 824 GAGAGAACATGATTAAG 765  
Qy 1229 CAATATTCGGGTGCAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1288  
Db 764 CAATATTCGGGTGCAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705  
Qy 1289 TGAAGTGAAGTGTGCTGAAGCTCAATGTTTGAAGTCTCAATTTGAGCACTTAATGAGA 1348  
Db 704 TGAAGTGAAGTGTGCTGAAGCTCAATGTTTGAAGTCTCAATTTGAGCACTTAATGAGA 645  
Qy 1349 ATTTCTGTGCAATGCTAGGTTTAATTTGAGACCAAAACATGTAAGACAGTATGAGATT 1408  
Db 644 ATTTCTGTGCAATGCTAGGTTTAATTTGAGACCAAAACATGTAAGACAGTATGAGATT 585  
Qy 1409 GAGTCAAGAAATTAAGATCAATGATGATGATGATGATGATGATGATGATGATGAT 1468

Db 584 GAGTCAAGATCTAGCATCATAGCTCCAGCTCCGCTGAGAGATGTGATCTCTCCAA 525  
Qy 1469 GGAAGAAAGAGAGAAACACCGGTTGTGGGCTGCACTGAGAGAAATAGAC- TGA 1527  
Db 524 GGAAGAAAGAGAGAAACACCGGTTGTGGGCTGCACTGAGAGAAATAGACCTTGA 465  
Qy 1528 AAGACGGCTCTCTAACATGTTTACAACTATCAACCTGTGA- TCAACGAGCACC 1586  
Db 464 AAGACGGCTCTCTAACATGTTTACAACTATCAACCTGTGATTCACGAGCACC 405  
Qy 1587 TTGTGACAGTTCTGTCCTTGTGTGATGACAAATTTT- GTGAAAAGTTTGTCAAT 1645  
Db 404 TTGTGACAGTTCTGTCCTTGTGTGATGACAAATTTTGTGAAAAGTTTGTCAAT 345  
Qy 1646 GTAGT- CAGAGTGCAGAAACCGCTT- CCGGAGTCCGCTGCAAGACAGTGCACAC 1703  
Db 344 GTAGTTCAGAGTGCAGAAACCGCTTCCCGGATCCGCTGCAAGACAGTGCACAC 285  
Qy 1704 CAAGCAGTCCCGCTGCTA- CCTGCTGTCCGAGAGTGAACCTGACCTGTCTTACTT 1762  
Db 284 CAAGCAGTCCCGCTGCTAACCCTGCTGTCCGAGAGTGAACCTGACCTGTCTTACTT 225  
Qy 1763 GTGAGCCGCTGACCATTTGGGACAGTAAATGTCTCTGCAAGAACTGAGTATTCAC 1822  
Db 224 GTGAGCCGCTGACCATTTGGGACAGTAAATGTCTCTGCAAGAACTGAGTATTCAC 165  
Qy 1823 GGGGCTCCAAAGACATGATGCTGGACCATCTGAGCT- GCGAGCTGGGGGATTTT 1881  
Db 164 GGGGCTCCAAAGACATGATGCTGGACCATCTGAGCTGAGCTGGGGGATTTT 105  
Qy 1882 ATCAAGATCTGTGCAAGAAATGATTT 1910  
Db 104 ATCAAGATCTGTGCAAGAAATGATTT 76

RESULT 8  
US-11-266-748A-98588/c  
Sequence 98588, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Hartkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcription Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 48396  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 98588  
LENGTH: 1140  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (86)..(86)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (169)..(169)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (362)..(362)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (367)..(367)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (468)..(468)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-98588

Query Match 43.8%; Score 980.8; DB 8; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 7.88-269;  
Matches 1074; Conservative 0; Mismatches 46; Indels 8; Gaps 6;

Qy 818 AGAAGAGCAAGCTTACATCTCTTCTATAGCTTTTCTGAGCGATGTTTAAATATG 877  
Db 1140 AGAAGAGCAAGCTTACATCTCTTCTATAGCTTTTCTGAGCGATGTTTAAATATG 1081  
Qy 878 ACTGCTTCATCATCTTTTCAATGACCAACCACTTAAAGCGAGAGCAAG 937  
Db 1080 ACTGCTTCATCATCTTTTCAATGACCAACCACTTAAAGCGAGAGCAAG 1021  
Qy 938 CAGCTTACCAACCACTTTGAGACCAAGTTTACAGATTTGAGGAGCAAG 997  
Db 1020 CAGCTTACCAACCACTTTGAGACCAAGTTTACAGATTTGAGGAGCAAG 961  
Qy 998 AGTTTCTGCTGCTCTACCGCTGAGGGGATTAAGACCCCAAGCTCCAGAGGCC 1057  
Db 960 AGTTTCTGCTGCTCTACCGCTGAGGGGATTAAGACCCCAAGCTCCAGAGGCC 901  
Qy 1058 GCAGAGGAGCGCTTCCCAATACAGTACAGGAGCCAGACCCCACTTATGTGC 1117  
Db 900 GCAGAGGAGCGCTTCCCAATACAGTACAGGAGCCAGACCCCACTTATGTGC 841  
Qy 1118 TGAATCAAGATACAGACAGTATAGGAGCGAGACTGAAACGGGGAGAGACA 1177  
Db 840 TGAATCAAGATACAGACAGTATAGGAGCGAGACTGAAACGGGGAGAGACA 781  
Qy 1178 ATGATTAAGAAAGAAAGAAAGAAATGAAATTTGAGCTCTTGAAGCAATCTC 1237  
Db 780 ATGATTAAGAAAGAAAGAAAGAAATGAAATTTGAGCTCTTGAAGCAATCTC 721  
Qy 1238 GGTGTCAACCAATTAAGTAAAGCAATATTTGAACCTCTGAGAAATGTGAGTGA 1297  
Db 720 GGTGTCAACCAATTAAGTAAAGCAATATTTGAACCTCTGAGAAATGTGAGTGA 661  
Qy 1298 GTGTGTGTAAGCTCAATGTTTGAAGTCTCATTTGCACTTACTATGACATTTCTGTG 1357  
Db 660 GTGTGTGTAAGCTCAATGTTTGAAGTCTCATTTGCACTTACTATGACATTTCTGTG 601  
Qy 1358 CCATTGCTAGTTTAAATGGGACCAAAACATGTACAGGTATGAGTTTGAAGTCAAG 1417  
Db 600 CCATTGCTAGTTTAAATGGGACCAAAACATGTACAGGTATGAGTTTGAAGTCAAG 541  
Qy 1418 AATCTAGCATATGCTCCAGCTCCGCTGAGATGTGATCTCTCAAGAAAAAGA 1477  
Db 540 AATCTAGCATATGCTCCAGCTCCGCTGAGATGTGATCTCTCAAGAAAAAGA 481  
Qy 1478 AGAGAAACAC- GATTGTGGCTGCACTGCAAGAAAGATACGTTGAAAAAGAGCGC 1536  
Db 480 AGAGAAACACGAGTTGTGGCTGCACTGCAAGAAAGATACGTTGAAAAAGAGCGC 421  
Qy 1537 TCCTTAACATGTTTACACTATCAACCTGTGATCATCAAGGAGCTTGT- GACA 1594

Db 420 TCCTCTAACCATGTTTACAGCATATGACCTCTGTGATCATCCAGCGAGCTTGATGACAA 361  
Qy 1595 GTTCGTGCCC-TTGTGTGATAGCAAAA--TTTTTGAAAAAGTTTGTCAATGAGTT 1651  
Db 360 GTTCGTGCCC-TTGTGTGATAGCAAAA--TTTTTGAAAAAGTTTGTCAATGAGTT 301  
Qy 1652 CAGAGTGTCAAAACCGCTTCCGGGATGCGGTGCAAGACAGTGCACACCAAGCACT 1711  
Db 300 CAGAGTGTCAAAACCGCTTCCGGGATGCGGTGCAAGACAGTGCACACCAAGCACT 241  
Qy 1712 GCCCGTGTACCTGTGCTGCCAGAGTGTGACCTGTGATCTGTCTTACTTGTGAGCGG 1771  
Db 240 GCCCGTGTACCTGTGCTGCCAGAGTGTGACCTGTGATCTGTCTTACTTGTGAGCGG 181  
Qy 1772 CTGACCATTTGGACAGTAAATATGTCTCTGCAAGAACTGCAAGTATTCAGCGGCTCA 1831  
Db 180 CTGACCATTTGGACAGTAAATATGTCTCTGCAAGAACTGCAAGTATTCAGCGGCTCC- 122  
Qy 1832 AAAGCATCTATTTCTGTGCAAGCATGTGAGTGCAGGCTGGGGGATTTTATCAAAGATC 1891  
Db 121 AAAGCATCTATTTCTGTGCAAGCATGTGAGTGCAGGCTGGGGGATTTTATCAAAGATC 63  
Qy 1892 CTGTGCAAAAATGAAATTCATCTCAGAAATCTGTGAGAGATTAATTT 1939  
Db 62 TCCTGGGCAAAAATGAAATTCATCTCAGAAATCTGTGAGAGATTAATTT 15

RESULT 9  
US-11-266-748A-151399  
Sequence 151399, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcription Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483396  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 151399  
LENGTH: 1140  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (673)..(673)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (774)..(774)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (779)..(779)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (972)..(972)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1055)..(1055)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-151399

Query Match 43.8%; Score 980.8; DB 8; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 7.8e-269;  
Matches 1074; Conservative 0; Mismatches 46; Indels 8; Gaps 6;

Qy 818 AGAGAGCAAAAGCTTACATCTCTTCAACGCTTTCTGAGGAGATGTTTAAATAG 877  
Db 1 AGAGAGCAAAAGCTTACATCTCTTCAACGCTTTCTGAGGAGATGTTTAAATAG 60  
Qy 878 ACTGCTTCTACATCTCTTCTTATGACACCAACCACTTATAGCGGAAGAACAGAAA 937  
Db 61 ACTGCTTCTACATCTCTTCTTATGACACCAACCACTTATAGCGGAAGAACAGAAA 120  
Qy 938 CAGCTGTGACAAACCTTGTGAGCAAGTGTACAGCATTTGAGGGAGCAAAAG 997  
Db 121 CAGCTGTGACAAACCTTGTGAGCAAGTGTACAGCATTTGAGGGAGCAAAAG 180  
Qy 998 AGTTGTCTGTCTCTCAACCGCTGAGCGATTAAGACCCCAACAAAGTCCAGAGAGCC 1057  
Db 181 AGTTGTCTGTCTCTCAACCGCTGAGCGATTAAGACCCCAACAAAGTCCAGAGAGCC 240  
Qy 1058 GCAAGAGAGACGCTTCCCAATTAACATAGAGGCCCAAGACCCCAACATTAATGTC 1117  
Db 241 GCAAGAGAGACGCTTCCCAATTAACATAGAGGCCCAAGACCCCAACATTAATGTC 300  
Qy 1118 TGAATCAAAAGATACAGACGTGTAGGAAGCAAGGAGCTGAACCGGGGAGAGAA 1177  
Db 301 TGAATCAAAAGATACAGACGTGTAGGAAGCAAGGAGCTGAACCGGGGAGAGAA 360  
Qy 1178 ATGATTAAG 1237  
Db 361 ATGATTAAG 420  
Qy 1238 GGTGTCAAAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297  
Db 421 GGTGTCAAAACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Qy 1298 GTGTGTGAGAGCGCTCAATGTTAGAGCTCATTTGCACTTACATGACAATTCGTG 1357  
Db 481 GTGTGTGAGAGCGCTCAATGTTAGAGCTCATTTGCACTTACATGACAATTCGTG 540  
Qy 1358 CCAATGTAGAGTTTATGGAGCAAAACATGTAGACAGGTATAGAGTTTATAGAGTCA 1417  
Db 541 CCAATGTAGAGTTTATGGAGCAAAACATGTAGACAGGTATAGAGTTTATAGAGTCA 600  
Qy 1418 AATCTAGCATATGCTCAGCTCCGCTGAGAGTGTGATATCTCTTCAAGAAAAAG 1477  
Db 601 AATCTAGCATATGCTCAGCTCCGCTGAGAGTGTGATATCTCTTCAAGAAAAAG 660  
Qy 1478 AGAGAAACACC-GTTGTGGCTGTGACACTGTGCAAGAAAGATACAGCTGAAGAGAGCGG 1536  
Db 661 AGAGAAACACC-GTTGTGGCTGTGACACTGTGCAAGAAAGATACAGCTGAAGAGAGCGG 720  
Qy 1537 TCCTCTAACATGTTTACATATCAACCTGTATCATCCAGCGAGCTTGT--GACA 1594  
Db 721 TCCTCTAACATGTTTACATATCAACCTGTATCATCCAGCGAGCTTGTGATGACAA 780  
Qy 1595 GTTCGTGCCC-TTGTGTGATAGCAAAA--TTTTTGAAAAAGTTTGTCAATGAGTT 1651  
Db 781 GTTCGTGCCC-TTGTGTGATAGCAAAA--TTTTTGAAAAAGTTTGTCAATGAGTT 840  
Qy 1652 CAGAGTGTCAAAACCGCTTCCGGGATGCGGTGCAAGACAGTGCACACCAAGCACT 1711







OY		352	CAGAAATTTTATGGTGGAAGAATAAAGCTGTTTTACATTACAATTCCTTTATATGCGAGATGAA	411
Dd		455	CAGAACCTTTATGTGTGAAGTAGAGACGGTTTTGTGCATATTTCCCTCAATGGAGATGMA	514
OY		412	GTTTTAGATCAGATGTACTCTTCATTGTAAGAACTAATAAAAAATTATGATGGAAAATA	471
Dd		515	GTGAAGAAGAAAGATGAGACTTTTATTATGAGAGCTGATCATTAACATATGATGGAAAGTc	574
OY		472	CACGGGGATAGAGA-----A-NGTGGTTTAAATGATGMAATTTTGTGNAG	519
Dd		575	CATGGTGAAGAGAGATGATGCCCTGGATCCGFTTCGATTATGTANGCTGTTTTCTGGAG	634
OY		520	TTCGTGAATGCCCCTTGTCATATTAATGATGATGCATGATG-----AT	564
Dd		635	TTGGTCGATGCCCCGTGAATCATGTA CTCGATGAGGAGAGAAAGGGCACATGACACCTCCA	694
OY		565	GATGAGACGATCCTGTAAGAAAAGAGAAAAAGCAAAGATCTGAGAGATCACCGAGAT	624
Dd		695	GATGAAAGCAGAGATGACGACAAAGAGATCTGCCAGTMACAAGAAAGAGAACGACAT	754
OY		625	GATPAAGAAACC GCC CCA CT CG ----- GAAATTT CCT TCG AT PA A ATT TT TG AA	675
Dd		755	GCTATTGAAGGCACAAAABAAGATTTCCAAGAAACAGTTCCCAAATGACATATCTTCAGT	814
OY		676	GCCATTTCTCAATGTGTTCCAGATTABGGSCACAGACAGAAAGACTAAAGAAAAATATAA	735
Dd		815	GCAATGTGCTCAATGTTCCCTGAGAAATGTGTGCCAAGATGACATGAAGAGAGATATCSA	874
OY		736	GAATCTACACCGAACAGACAGCTCCCAAGGGCACTTTCCTCGTAATGTACCCCCCAATAGAT	795
Dd		875	GAACCTAACAGAGATGTCAGACCCCAATCACCTTCCCTCAGTGCACCCCAACATGCAT	934
OY		796	GAACCAATGCTAAATCTGTTCAAGAGAGCAAAGCTACACTCCTTCATACGCTTTTC	855
Dd		935	GGCCCCAATGCAAGTCTGTGACGGGGAGCAATCTCGACCTCCTTCCACACACTTTT	994
OY		856	TGTAGGGGATGTTTAAATATGACTGCTTTCCTACATCTTTTCATGCAACCCCAACT	915
Dd		995	TGCCGGCCCTCTTTAAATAGACAGCTCTCTTCAACCTTTTCATGCAACCCCTTAATGTA	1054
OY		916	TATAAGGGGAAGACACAGAAACAGCTCAGACAA CA AACCTTGAGCA C AG GTT AC	975
Dd		1055	TATTAACCGAAGATTAAGAAANTCAAGATTGAACAGAACCAATGTSGACAGACTGCTTC	1114
OY		976	CAGCATTTGAGGGAGCAAAGAGATTGCTGCTCTCAGCCGTGACGGATTAAGACC	1035
Dd		1115	CTTTTGTGGAAGGAGCAAAGAGATAT-----GCCATGCTCCACAACACCC	1159
OY		1036	CCAACBAACGTCCAGAGAGCCCGCAGAAAGACGGCTTCCCAATTAACATAGACAGGCC	1095
Dd		1160	CGCTCCAAAGTGTCTGATCGTGC CG GAAGAGCCACCAATAGTCA GTGCTTCGCTCC	1219
OY		1096	AGCAACCCCAACATTAATGTGCTGGAATCAAAAGATACAGACAGTGAATAGGAAGACAGGG	1155
Dd		1220	AATGECTAGCCTCTGCTGTGGCTGAGCTTAAGAGAGAGACAGTGAACGGGACACAGGC	1279
OY		1156	ACTGAACCGGGGGAGAGAACATATGATAAGAGAGAAAGAGAAAGATGAATCTTG	1215
Dd		1280	AATG-----ACTGGGCTCCACAGTTCTTCAGAGGCTMACTCTCGCTGTCAg	1324
OY		1216	AGCTCCTGTGAAGCAATTC TC GG GTG TA A C ACC ATA AAA GAT GA A GC AAA AT TT GA A	1275
Dd		1335	ACTCCACAAACAGAAAGGCT---AGTCAGCCCCCACTCAACTCTGCGTATGTGAACSA	1381
OY		1276	CCTCCTGAAGATGTGAGTGAAGTGTGCTGAAGCTCAATGTTTTAGATCCTCATTTGGC	1335
Dd		1382	CCCTCGAGGCGTGTGAATGAGCTGGGGCTGAAGATCTCTTTTTCAGTCTTCCATGCG	1441
OY		1336	ACTTACTATGACATTTCTGTGCCATTTGCTATGTTAATTTGGACCAAACAATGTAACAG	1395
Dd		1442	ACCTACTTCAACACCTCTCTTTCATATACACAGGCTTCGGGGACCAACACGTGCAMAGAG	1501

QY	1396	GTGATGATGATTTAAGTCAAAAGATCTAGATCATAGCTCAGTCCCGGTGAGAAATGTG	1455
Db	1502	GTCTTTACGTTTGACGTCAAAAGATCACTTATCTGAACTGCGCAACAGATGAGCTCATG	1561
QY	1456	GATATCTCTCCAAAGAAAAAGAAAGAGAAACACCGGTTGTGGGTGCACATCGACAAAG	1515
Db	1562	---AACCCTCAGAAAGAAAGAAAAAGAAAGAACAGATGTGTGGGTGCACATCGACAAAG	1618
QY	1516	ATPACGCTGAAAAAGACGGCTCCTCTTAACCATGTTTACATATCAATCACTTGATCAT	1575
Db	1619	ATTGACGCTGAAGAAAGATTAATCTTTCACACAAGGTGTACAACTACCAACCTCGACAC	1678
QY	1576	CCACGGGACCTTGTGACAGTTGCGTGCCTTGTGTGATACACAAATTTTGTGAAAG	1635
Db	1679	CCAAACGGCCCTGTGACAGACCTGCCCCCTGCATCATGTACAGAAATTTCTGTGAAAG	1738
QY	1636	TTTTGTCAATATGATTGACAGGTGTCAAAACCGCTTTTCCGGGATGCGCTGCAAGCAC	1695
Db	1739	TTCTGCGAGTGCACACCGAGCTGTCAAAATGTGTTTCCCTGTGCTGTGTAAGACCCAG	1798
QY	1696	TGCAACACCAAGCAGTSCCGGTGTACCTGTGTCCGAGATGTGACCTGTACCTGTGT	1755
Db	1799	TGCAATATCAAGCATGTGCTTGTGTAATCTGCGAGTGTGGAATGTGACCTGTGCTGT	1858
QY	1756	CTTACTTGTGAGGCCGTGTGACCATTTGGGACAGTAAAAATGTGTCTCGCAAGAACTGAGT	1815
Db	1859	CTCACCTGTGTGGGGCTTCAGAGCACCTGTGGACCTGCAGAGGTGTTCCTGTAAAACTGAGC	1918
QY	1816	ATTCAGCGGGGCTCCAAAAAGCATCTATGTGTGACACATCTGACGTGTGACGCTGGGGG	1875
Db	1919	ATCCAGCTGTGACTTTAAGAGACACTGTGTGCCCCCTCTGTATGTGTGCGGATGGGGC	1978
QY	1876	ATTTTATCAAAAGATCTGTGTGACAGAAAAATGAATTCATCTCAGAAATCTGTGAGAGATT	1935
Db	1979	ACCTTCATAAAGAGGTGTGTGACAGAAAGCAATTCATTTCTGAATCTGTGTGAGCTC	2038
QY	1936	ATTTCTCAAGATGAAGTGTGACAGAAAGGGAAGTGTATGTATTAATTCATGTGACGTTT	1995
Db	2039	ATCTCTCAGATGTAGGCTGTGATCGACGGGAAAGGTCTATGACAAATTCATGTCCAGCTTC	2098
QY	1996	CTGTTCACCTTGAAACAATGATTTTGTGTGTGATCAACCCGACAGGGGTAAACAAATTCGT	2055
Db	2099	CTCTTCACCTCAATTAATGATTTTGTGTGTGATCTGTGTGGAAGGAACAAATTCGA	2158
QY	2056	TTTGCAAATCATTCGGTAAATCCAACTGTCTATGCAAAAGTATGATGTTAACGTGAT	2115
Db	2159	TTTGCAAATCATTCAGTGAATCCCACTGTATGCCAAAGTGTGATGTAATGAGGAC	2218
QY	2116	CACAGGATAGGATTTTGTGCCAAGAGGCCATCCAGACTGGCCGAAGGCTGTTTTTGAT	2175
Db	2219	CATCGGATTTGGGATCTTTGGCCAAAGGGGCAATTCATACCTGGCCGAAGGCTTCTTTGAT	2278
QY	2176	TACGATTCACCGCAGGCTGATGCGCTGTAAGATGTGCGGATCGAAAGGAATGGAATTC	2235
Db	2279	TACAGGTACAGCCCAAGCTGATGCTCTCAAGTACGTGGGATCGAAGGAGACCGACGTC	2338
QY	2236	CCTT 2239	
Db	2339	CTTT 2342	
RESULT 13			
US-11-266-748A-26094			
Sequence 26094, Application US/11266748A			
Publication NO. US20060134663A1			
GENERAL INFORMATION:			
APPLICANT: Harkin, Paul			
APPLICANT: Johnston, Patrick			
APPLICANT: Mulligan, Karl			
TITLE OF INVENTION: Transcriptome Microarray Technology and			
TITLE OF INVENTION: Methods of Using the Same			
FILE REFERENCE: 55815-0102 (319189)			
CURRENT APPLICATION NUMBER: US/11/266,748A			

```

: CURRENT FILING DATE: 2005-11-03
: PRIOR APPLICATION NUMBER: EP 04105479.2
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105482.6
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105483.4
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105507.0
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105485.9
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105484.2
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: US 60/662,276
: PRIOR FILING DATE: 2005-03-14
: PRIOR APPLICATION NUMBER: US 60/700,293
: PRIOR FILING DATE: 2005-07-18
: NUMBER OF SEQ ID NOS: 483996
: SOFTWARE: PatentIn version 3.3
: SEQ ID NO 26094
: LENGTH: 4640
: TYPE: DNA
: ORGANISM: Homo Sapiens
: US-11-266-748A-26094

```

Query Match	37.5%	Score 840.4	DB 8	Length 4640
Best Local Similarity	64.2%	Pred. No. 1.8e-226		
Matches 1428	Conservative	0	Mismatches 721	Indels 75
			Gaps	8
Qy	55	GTAAATCAGAGTACATCGACTGAGCAGCTCAGAGGTTTCAGACGAGCTGATGAATTA	114	
Db	177	GTGAATCTGATACATCGACTGAGCAGCTTGAACAATGCGCTTCAGGCAATATGGGTCA	236	
Qy	115	AAGGATGTTTAACTTCATCGTCAGAAAATTTTGAAGAAGCGAAATCTTAAACCA	174	
Db	237	AAGGCTTTGATGTTGGCAAAATTTTCAAGGTTCAAGAAAACCCAGATCCTCAATGA	296	
Qy	175	GAATGGAACACCGGAGATACAGCTGTGCACATCTGACATCTTGTAGCTCATTTGGC	234	
Db	297	GAATGGAAGAAGCTTCGTGCCAACCTGTCAGTCAATGAAGCCGTGAGTGAACACCT	356	
Qy	235	GGGACTAGAGGTGTC---GGTGCACAGTGACTTGGAATTTTCCAAACAAGTCATCCCA	291	
Db	357	TTTCTCAAAAAGTGAACCATAGAGGCAATTTTCCGGGATTTTGCAGCCCAATATGTTA	416	
Qy	292	TTAAAGCTCGAATCGAGTGTCTTCAGTACCCATATGATCTTGTGCTCCCTTAG	351	
Db	417	ATGAGGTCACTGAACACAGTTGCAATGTTGCCATCATGATATCTGTGTCCTTCCTCA	476	
Qy	352	CAGAAATTTTATGTGGAAGATGAACCTGTTTACATTAATCTTATATGAGAGATGA	411	
Db	477	CAGAACTTTATGTGTGAAGATGAGCGGTTTGTGCAATATCTTCAATGAGAGATGA	536	
Qy	412	GTTTATGATCAGAGTGTACTTTCATTGGAAGACTAATATATATATATATATGAGAACTA	471	
Db	537	GTTGAAGAAGAGATGAGACTTTTATGAGAGCGTCAATATCAATGATGAGGAAATTC	596	
Qy	472	CACGGGGATAGGA-----ATGTGGTTTATTAATGATGAATTTTGTGAG	519	
Db	597	CATGTGTAAGAGAGATGATCCCTGGAATCCGTTCTGATTTAGTATGCTGTTTTCTGAG	656	
Qy	520	TTGTGATATGCCCTTGTGTCATATATATATGATGATGATGATG-----AT	564	
Db	657	TTGTGATATGCCCTTGTGTCATATATATGATGATGATGATGATGATGATGATGATGATG	716	
Qy	565	GATGAGACGATCTTGAAGAAAGAGAAAGAAAGCAAGAAAGATCTGAGAGATCACCGAAT	624	
Db	717	GATGAGAAAGCAGAGATGACAGCAAGAAAGATCTGCAATACAGAAAGAAAGCAAGAT	776	
Qy	625	GATTAAGAAAGCCGCCACCTCG-----GAAATTCCTTCGATTAATTTTGA	675	
Db	777	GCTATTTAAGGCAAAAAGAGATTTCAAGAAACGTTTCCCAATATGACATGATCTTCAGT	836	

QY	676	GCATTTCTCAATGTTTCCAGATTAAGGGCACAGCAAGAAAGCTAAAGAAAATATATAA	735
Db	837	GCAATGCTCAATGTTCCCTGAAGAAATGTGTCCCAATGACATGAAGAAAGGTATGCA	896
QY	736	GAATCAACCGAACAGACAGCTCCCAAGCGCACTTCTCTGATATATACCCCAACATAGAT	795
Db	897	GAATCAACAGAGATGTCAAGACCCCAATGCACTTCCCTCAGTGCACCCCAACATGAT	956
QY	796	GAACCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACACTCTTTCAATACGCTTTTC	855
Db	957	GGCCCAATGCAAGCTGTGTCAACGGGAGCAATCTGTGCACTCTCTTCCACACTTTT	1016
QY	856	TGTAGGCGAATGTTTAAATATAGCTGTCTTCAATCTTTCAATGCAACCCAACT	915
Db	1017	TGCGGCGCTCTTAAATATACAGCTGCTCTTCACTTTTCATGCAACCCCTTAATGTA	1076
QY	916	TATAGCGGAAGAAACACAGAAACAGCTCTAGACAAACCTTGTGACCAAGTGTAC	975
Db	1077	TATTAACCGCAAGAAATTAAGAAATCAAGATTGACCAAGAACATGTGGACAGACTGCTTC	1136
QY	976	CAGCATTTGAGGAGCAAAAGAGTTTGTGTGCTGTCTCACCGCTGACGGATTAAGACC	1035
Db	1137	CTTTTGCTGGAAGAGCAAAAGAGTAT-----GCCATGCTCCACAAACCC	1181
QY	1036	CCACCAAAACGTCCAGAGGCGCCGACAGAGAGAGAGCGCTTCCCATTAACATAGACAGGCC	1099
Db	1182	CGCTCAAGTGTCTGT	1241
QY	1096	AGCAACCCCAACCATTAATGTGTGTAATCAAAAGATACAGACAGTATAGAGGAAGCAGG	1155
Db	1242	AATGCTTACGCTCTGT	1300
QY	1156	ACTGAAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1215
Db	1302	AATG-----ACTGGGCTCCAGTTCCTTCAAGAGGCTTAATCTGTGTGTGTGTGTGTGTGTGT	1346
QY	1216	AGCTCTCTGAGCAAAATTTCTGGTGTCAACACCATTAAGATGAGGCCAAATATTGAA	1275
Db	1347	ACTCCCAAAACAGAAAGGCT--AGTCAAGCCCACTCACTGTGTGTGTGTGTGTGTGTGTGT	1403
QY	1276	CCTCTGAGAAATGAGAGTGGAGTGGGTGAGAGCTCAATGTTAGAGTCTGATTTGAC	1333
Db	1404	CCCTGGAGCTGTGGAATGAGCTGGGGCTTAAGAAATCTTTTTCAGATCTTTCATATGAC	1466
QY	1336	ACTTACTATGCAAAATTTCTGTGCCATTGTAGTTAAATTTGGACCAAAACATGTAGACAG	1395
Db	1464	ACCTACTTCAACAATCTGTCTTCAATAGCCAGGCTTGTGGGACCAAGACGTGCAAGACAG	1522
QY	1396	GTTATATAGATTTAAGTCAAAAGATCTAGCATCATAGCTCAAGCTCCCGCTGAGAGATGTG	1455
Db	1524	GTTCTTCAATTGAGCTCAAAAGATCTTATCTTGAAGCTGCCAAGCAGATGATCACTCATG	1583
QY	1456	GATATCTCTCCAAAGAAAAGAAAGAAACACCGTTGTGGGCTGCAACATGAGAAAG	1515
Db	1584	---AACCCCTCACAGAAAGAAAGAAAGAAAGCAAGATTTGTGGCTGCAACTGTGAGAG	1640
QY	1516	ATAAGCTGAAAAGAGACGGCTCTCTTCAACCATGTTTACAATTAACAACCTGTGATCAT	1575
Db	1641	ATTATGCTGAAGAAAGATTAATCTTTTCCACACAAATGTATCAACTACAAACCTGTGACCA	1700
QY	1576	CCACGGACGCTTGTGACAGT	1635
Db	1701	CCAGACGGCCCTGTGACAGCACTGTGCCCTGTGACATCATGATCAGAAATTTCTGTGAGAG	1766
QY	1636	TTTTTGTCAATGTAATTTCAAGTGTCAAAACCGCTTTCGGGATGCCGTGCAAGACAG	1699
Db	1761	TTTGTGCAAGTGAACCAAGACTGTCAAAATGTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT	1820
QY	1696	TGCAACACCAAGACAGTCCCGTGTGATCTGTGCTGTGCGAGAGTGTGACCTGTGACTGTGT	1755
Db	1821	TGCAATACCAAGACATGTCTTGT	1880
QY	1756	CTTACTTGTGAGCGCTGACCAATTTGGGACATTAATAATGTGTCTTCCAGACATCTGACT	1815

Db 1881 CTGACCTGTGGGGCTCAGAGCACTGGGACTGCAAGGTGTCTCTGTAAAACTGACGC 1940  
Qy 1816 ATTGAGGGGGCTCCAAAAGCATCTATTGCTGGACCATCTGACGTGGCGGGCGG 1875  
Db 1941 ATTCAGGCTGGACTTAAAGACACTGCTGCGCCCCCTGTGATGTGGCGGAGTGGC 2000  
Qy 1876 ATTTTATCAAGATCTGTGCAAGAAAATGAAATTCATCTGCAATATCTGTGAGAGATT 1935  
Db 2001 ACCCTTCAATAAAGAGTCTGTGCAAGAGACGAAATTCATTTCTGAATATCTGTGGAGCTC 2060  
Qy 1936 ATTTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTATGATTAATATCATGTGACCTTT 1995  
Db 2061 ATCTCTGAGATGAGGCTGATGACGCGAAAGGCTATGACAAATATCATGTCCAGCTTC 2120  
Qy 1996 CTGTTCAACTGGAACATGATTTTGTGTGTGATGCAACCCCAAGGGTAAACAAATTCGT 2055  
Db 2121 CTCTTCAACCTCAATTAATGATTTGTATGTGATGCTACTCGAAAGGAAACAAATTCGA 2180  
Qy 2056 TTTCGAATCATTCGGTAAATCCAAACTGCTATGCAAAAGTTATGATGTTAACGGTGTAT 2115  
Db 2181 TTTCGAATCATTCAGTGAATCCCACTGTTATGCAAAAGTGTATGTGATGAGAGAC 2240  
Qy 2116 CACAGGATAGGTATTTTGGCCAGAGAGCCATCCAGACTGGCGAAAGACTGTTTGTAT 2175  
Db 2241 CATGGATTTGGGATCTTTGCCAAGAGGGCAATTCAAAGCTGGCGAAAGACTCTTCTTTGAT 2300  
Qy 2176 TACGATTCAGCCGCGGTGATGCCCTGAAGTATGTGCGGATCGAAGAGAAATGGAATC 2235  
Db 2301 TACAGGTACAGCCAGAGTGAATGCTCTCAAGTACGTGGGATCGAGAGGAAACCGACCTC 2360  
Qy 2236 CCTT 2239  
Db 2361 CTTT 2364

RESULT 14  
US-11-266-748A-175796  
Sequence 175796, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 48396  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 175796  
LENGTH: 1011  
TYPR: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc\_feature

/ LOCATION: (18)..(18)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (64)..(87)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (141)..(141)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (151)..(151)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (212)..(212)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (222)..(222)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (238)..(238)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (251)..(251)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-175796

Query Match 37.4%; Score 837.2; DB 8; Length 1011;  
Best Local Similarity 93.2%; Pred. No. 6.1e-228;  
Matches 935; Conservative 0; Mismatches 57; Indels 11; Gaps 7;

Qy 483 AGATGTGGGTTTAAATGATGAAATTTTGTGGAATTTGATGAAATGCTTGTGTC-AAT 541  
Db 1 AGAATGTGGGTTTAAATGATGAAATTTTGTGGAATTTGATGAAATGCTTGTGTC-AAT 60  
Qy 542 AATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 601  
Db 61 AATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
Qy 602 AAGATGTGAGGATCACCGAG--ATGATAAAGAAAGCCGCCACCTCGAAATTTCTTC 659  
Db 121 AAGATGTGAGGATCACCGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
Qy 660 T--GATTAATTTTGAAGCATTTCTCAA--TGTTTCAAGTAAAGGACACGCA-GA 713  
Db 181 TGTGATAAATTTTGTGAAGCATTTCTCAATGTTTCCAMATGTTTCCAMATGTTTCCAMAT 240  
Qy 714 AGAATAAG-GAAAAATATTAAGAACTCAACCAACAGAGCTCCAGGCGCATTTCTCTC 772  
Db 241 AGAATAAGGAAAAATATTAAGAACTCAACCAACAGAGCTCCAGGCGCATTTCTCTC 300  
Qy 773 CTGATGATCCCCCAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 832  
Db 301 CTGATGATCCCCCAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
Qy 833 TACACTCTTTTCAATAGCTTTTCTGTAGGGAATTTTAAATAGATGATGATGATGATGATGATGAT 892  
Db 361 TACACTCTTTTCAATAGCTTTTCTGTAGGGAATTTTAAATAGATGATGATGATGATGATGATGAT 420  
Qy 893 CTTTTCATGCAAGCCCAACTTATAGGGAAGAAACAGAAACAGCTCTAGACAA 952  
Db 421 CTTTTCATGCAAGCCCAACTTATAGGGAAGAAACAGAAACAGCTCTAGACAA 480  
Qy 953 AACCTTGTGACCAAGTGTACAGATTTTGAAGGAGCAAAAGAGTTTGTCTGTCTC 1012  
Db 481 AACCTTGTGACCAAGTGTACAGATTTTGAAGGAGCAAAAGAGTTTGTCTGTCTC 540  
Qy 1013 TCACGCTGAGCGGATTAAGACCCCAACCAAAAGCTCCAGAGAGCGGACGAAAGAGAGCGGC 1072

Db 541 TCACGCTGAGGCGATTAAGACCCCAACCAAAAGTCAGAGAGGCGCAGAGAGCGGC 600  
Qy 1073 TTCCCAATACAGTAGACAGCCCAACCAACCAATTAATGCTGGAATCAAGAGATA 1132  
Db 601 TTCCCAATACAGTAGACAGCCCAACCAACCAATTAATGCTGGAATCAAGAGATA 660  
Qy 1133 CAGACAGTAGAGGAG 1192  
Db 661 CAGACAGTAGAGGAG 720  
Qy 1193 AAG 1252  
Db 721 AAG 780  
Qy 1253 TAAAGATGAG 1312  
Db 781 TAAAGATGAG 840  
Qy 1313 CAATGTTTGAAGTCCCTCAATGAGCACTTAATGAGCAATTTCTGTCGCAATGCTTAA 1372  
Db 841 CAATGTTTGAAGTCCCTCAATGAGCACTTAATGAGCAATTTCTGTCGCAATGCTTAA 900  
Qy 1373 TTGGGACCAAAACATGTAGACAGGTGTATGAGTT--TAGAGTCAAGAATCTAGCATCATA 1431  
Db 901 TTGGGACCAAAACATGTAGACAGGTGTATGAGTTAGGTCAAGAATCTAGCATCATA 960  
Qy 1432 GCTCCAGCTCCGCTGAGAGATGTGATCTCTCCAGAGAGAA 1474  
Db 961 GCTCCAGCTCCGCTGAGAGATGTGATCTCTCCAGAGAGAA 1003

RESULT 15  
US-11-266-748A-361225/c  
; Sequence 361225, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcription Microarray Technology and  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; CURRENT FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: Patencin version 3.3  
; SEQ ID NO 361225  
; LENGTH: 904  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; PEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (30)..(30)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature

; LOCATION: (134)..(134)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (139)..(139)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (205)..(205)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (240)..(240)  
; OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-361225

Query Match 34.2%; Score 765.6; DB 8; Length 904;  
Best Local Similarity 98.4%; Pred. No. 1.4e-207;  
Matches 814; Conservative 0; Mismatches 7; Indels 6; Gaps 4;

Qy 826 CAAGCTTACACTCTCTTCATACGCTTTCTGTAGGCGATGTTTAATATGACTGCTTC 885  
Db 904 CAAGCTTACACTCTCTTCATACGCTTTCTGTAGGCGATGTTTAATATGACTGCTTC 845  
Qy 886 CTACATCTCTTCATGCAACCCCAACCTTATAGGGAGAAACACAGAAACAGCTCTA 945  
Db 844 CTACATCTCTTCATGCAACCCCAACCTTATAGGGAGAAACACAGAAACAGCTCTA 785  
Qy 946 GACAAACAACCTTGTGACACACAGTGTACGACATTTGAGAGAGAGAGAGAGAGAG 1005  
Db 784 GACAAACAACCTTGTGACACACAGTGTACGACATTTGAGAGAGAGAGAGAGAGAG 725  
Qy 1006 GCTGCTTCACCGCTGAGCGGATTAAGACCCCAACCAACCTCCAGAGAGCGCGAGAGA 1065  
Db 724 GCTGCTTCACCGCTGAGCGGATTAAGACCCCAACCAACCTCCAGAGAGCGCGAGAGA 665  
Qy 1066 GGAAGGCTCCCAATTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1125  
Db 664 GGAAGGCTCCCAATTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605  
Qy 1126 AAGATACAGACAGTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1185  
Db 604 AAGATACAGACAGTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545  
Qy 1186 GAAAG 1245  
Db 544 GAAAG 485  
Qy 1246 ACACCAATTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305  
Db 484 ACACCAATTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
Qy 1306 GAAAGCTCAATGTTTGAAGTCTCTATGAGCACTTAATGAGCAATTTCTGTGCAATTC 1365  
Db 424 GAAAGCTCAATGTTTGAAGTCTCTATGAGCACTTAATGAGCAATTTCTGTGCAATTC 365  
Qy 1366 AGTTTAATGAG 1425  
Db 364 AGTTTAATGAG 305  
Qy 1426 ATCATAGCTCAGAGCTCCGCTGAGAGATGTGATCTCTCCAGAGAGAGAGAGAGAG 1485  
Db 304 ATCATAGCTCAGAGCTCCGCTGAGAGATGTGATCTCTCCAGAGAGAGAGAGAGAG 245  
Qy 1486 CACC-GGTTGGGCTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1544  
Db 244 CACAGGTTGGGCTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185  
Qy 1545 CCATGTTTACATATCAACCTCTGATCATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1602  
Db 184 CCATGTTTACATATCAACCTCTGATCATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 125  
Qy 1603 CC-TTGTGTAGTAGCAAAA--TTTGTGTAAAAGTTTGTCAATG 1646

Db 124 CCTTGTGTGATGACACAAAATTTTGTGTAAGTTTGTCAATG 78

Search completed: August 13, 2006, 17:06:06  
Job time : 529 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:16:19 ; Search time 131.576 Seconds  
(without alignments)  
2428.962 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGQYKSKSEKGPVCMKRRVK.....KGNKIRFANHSVNPNCYAKV 699

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_8:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*
- 10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3808	100.0	746	2	AAW05260
2	3802	99.8	746	6	ABU56461 Lung canc
3	3802	99.8	746	7	ADG10642 Human STA
4	3802	99.8	746	7	ADN39392 Cancer/an
5	3787.5	99.5	751	8	ADP95920 Human E2H
6	3569.5	93.7	707	7	ADP65008 Human pro
7	2802	68.3	517	7	ADG10764 Human STA
8	2384.5	58.1	723	6	ADX05848 Cyclin-de
9	2212.5	52.6	747	9	ADU11057 Human pro
10	1959.5	51.5	760	4	ABB62427 Drosophila
11	719.5	18.9	880	8	ADY25049 Plant ful
12	717	18.8	895	8	ADT57579 Plant pol
13	698	18.3	893	5	AAO14530 Protein o
14	693	18.2	891	5	ADM48181 Polypepti
15	677	17.8	931	8	ADT57263 Plant pol
16	665	17.5	921	3	AAV32297 Corn poly
17	663	17.4	931	5	AAO14529 Protein o
18	619	16.3	509	8	ADX94792 Plant ful
19	612	16.1	689	2	AAV42698 Arabidops
20	612	16.1	689	4	AAV57036 Feritilisa
21	612	16.1	689	4	AAAB82455 Arabidops
22	612	16.1	689	6	ABP70523 Histone d
23	612	16.1	932	6	ABP70537 Amino aci

24	611	16.0	689	3	AAB01673
25	607.5	16.0	463	8	ADX94788
26	601	15.8	417	8	ADX95305
27	582.5	15.3	536	8	ADY13472
28	569.5	15.0	346	8	ADX94796
29	547	14.4	294	4	ADY14004
30	528.5	13.9	326	8	ADX66894
31	525.5	13.8	298	8	ADX93483
32	510	13.4	256	8	ADY71472
33	504.5	13.2	244	8	ADX94793
34	470.5	12.4	225	8	ADX94797
35	470.5	12.4	359	3	AAV32298
36	451	11.8	196	9	ADY65574
37	409	10.7	102	4	AAU15631
38	409	10.7	102	4	ABB34636
39	409	10.7	102	4	AAU28129
40	409	10.7	102	4	ABB29459
41	409	10.7	102	4	AAU67821
42	409	10.7	102	4	AAU03374
43	409	10.7	102	5	ABG37364
44	318	8.4	153	8	ADY71195
45	293	7.7	114	10	AEE98254

#### ALIGNMENTS

AAB01673 F1S1 prot  
Adx94788 Plant ful  
Adx95305 Plant ful  
Ady13472 Plant ful  
Adx94796 Plant ful  
Ady14004 Plant ful  
Adx66894 Plant ful  
Adx93483 Plant ful  
Adx71472 Plant ful  
Adx94793 Plant ful  
Adx94797 Plant ful  
Aay32298 Wheat pol  
Ady65574 S. manson  
Aam15631 Peptide #  
Abb34636 Peptide #  
Aam28129 Peptide #  
Abb29459 Peptide #  
Aam67821 Human bon  
Aam03374 Peptide #  
Abg37364 Human pep  
Ady71195 Plant ful  
Aee98254 Human E2H

RESULT 1  
ID AAW05260 standard; protein; 746 AA.  
XX  
AC AAW05260;  
XX  
DT 25-MAR-2003 (revised)  
DT 05-MAY-1997 (first entry)  
XX  
DE Chromatin regulator protein E2H2.  
XX  
KW Chromatin; regulator; E2H1; E2H2; SUV39H; SET domain; gene therapy;  
KW cancer.  
XX  
OS Homo sapiens.  
XX  
PN W09635784-A2.  
XX  
PD 14-NOV-1996.  
XX  
PF 02-MAY-1996; 96WO-BP001818.  
XX  
PR 10-MAY-1995; 95DB-01016776.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.  
XX  
PI Jenuwein T, Laible G;  
XX  
DR WP1; 1996-518672/51.  
XX  
PT N-PSDB; AAT43624.  
XX  
PT New DNA encoding chromatin regulator protein with SET domain - and  
XX related vectors, transformed cells, proteins and antibodies, for  
XX diagnosis and treatment of cancer.  
XX  
PS Claim 12; Fig 7; 38pp; German.  
XX  
CC The DNA was isolated by screening a human B cell cDNA library with mixed  
CC Drosophila DNA probes based on the conserved SET domains in B(2) and  
CC Suvar(3-9). The DNA, and its products, are useful in therapy (esp. gene  
CC therapy) and diagnosis of human diseases that involve deregulated  
CC chromatin-regulator genes having a SET domain, esp. cancer. (Updated on  
CC 25-MAR-2003 to correct PR field.)  
XX  
SQ Sequence 746 AA;

Query Match 100.0%; Score 3808; DB 2; Length 746;  
Best Local Similarity 100.0%; Pred. No. 6,9e-282;  
Matches 699; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MGQTKSEKGPVCMRKRVKSEYMRRLQKRFPRADSVKSMFSSNRQKILERTIINQEW 60
DB 1 MGQTKSEKGPVCMRKRVKSEYMRRLQKRFPRADSVKSMFSSNRQKILERTIINQEW 60
QY 61 KORIQPHILITVSISLGRTRGCVTSDDLPPPOVIFLKTLANAVASPIYMSPLQONF 120
DB 61 KORIQPHILITVSISLGRTRGCVTSDDLPPPOVIFLKTLANAVASPIYMSPLQONF 120
QY 121 MVEDETVLHNIPIYMGDEVLDQDGFIFIELIKNDGKVGHRECGFINDEIFVELVNALGQ 180
DB 121 MVEDETVLHNIPIYMGDEVLDQDGFIFIELIKNDGKVGHRECGFINDEIFVELVNALGQ 180
QY 181 YNDDDDDDGDDPPEEREKQDLSDHDDKESRPRKPSDKITFEALISSMFPDGTABEL 240
DB 181 YNDDDDDDGDDPPEEREKQDLSDHDDKESRPRKPSDKITFEALISSMFPDGTABEL 240
QY 241 KEKTKELTEOOLPQALPPECTPNIDGNPAKSVQREOSLASHFHTLFCRCRCFYDCFLHPFH 300
DB 241 KEKTKELTEOOLPQALPPECTPNIDGNPAKSVQREOSLASHFHTLFCRCRCFYDCFLHPFH 300
QY 301 ATPNTYRKNTETALDNKPCGPOCYOHLGAKSPAAALTARITKTPPKRPGRRGRRLPN 360
DB 301 ATPNTYRKNTETALDNKPCGPOCYOHLGAKSPAAALTARITKTPPKRPGRRGRRLPN 360
QY 361 NSSRPSPTINIVLSEKOTDSDEAGTENGENNKEBEKEDENSSSEANSRCOTPIKM 420
DB 361 NSSRPSPTINIVLSEKOTDSDEAGTENGENNKEBEKEDENSSSEANSRCOTPIKM 420
QY 421 KPNIEPPENVMSGAEASMFVILIGTYVDNFCALIRLIGTYCQVVEFRVKESSIIAPA 480
DB 421 KPNIEPPENVMSGAEASMFVILIGTYVDNFCALIRLIGTYCQVVEFRVKESSIIAPA 480
QY 481 PAEDVTPPRKKRKRIMAAHCKRIQLKDGSSNHTVNYPCCHPROPCSSGCPVYAAQ 540
DB 481 PAEDVTPPRKKRKRIMAAHCKRIQLKDGSSNHTVNYPCCHPROPCSSGCPVYAAQ 540
QY 541 NPCKPQCCSSECONRPRGCRCKAQCNKQCPCTLAVERCPDILCTCGAADHMDSKNVS 600
DB 541 NPCKPQCCSSECONRPRGCRCKAQCNKQCPCTLAVERCPDILCTCGAADHMDSKNVS 600
QY 601 CKNCISIQSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGRIISQDEADRRGKYVDK 660
DB 601 CKNCISIQSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGRIISQDEADRRGKYVDK 660
QY 661 YMCSPFLPNLNDPVVDATRKANKIRFANHVSVPNCYAKV 699
DB 661 YMCSPFLPNLNDPVVDATRKANKIRFANHVSVPNCYAKV 699
QY 699 YMCSPFLPNLNDPVVDATRKANKIRFANHVSVPNCYAKV 699
DB 699 YMCSPFLPNLNDPVVDATRKANKIRFANHVSVPNCYAKV 699
RESULT 2
ABUS6461
ID ABUS6461 standard; protein; 746 AA.
AC ABUS6461;
XX 02-APR-2003 (first entry)
DB Lung cancer-associated polypeptide #54.
XX Lung cancer-associated polypeptide; cytosolic; emphysema;
XX antiinflammatory; antisthmatic; non-small cell lung cancer; atelectasis;
XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX unidentified.
XX OS
XX FN WO200286443-A2.
```

```
PD 31-OCT-2002.
XX 18-APR-2002; 2002MO-US012476.
PF 18-APR-2002; 2002MO-US012476.
XX 18-APR-2001; 2001US-0284770P.
PR 10-MAY-2001; 2001US-0290492P.
PR 09-NOV-2001; 2001US-0339245P.
PR 13-NOV-2001; 2001US-0350666P.
PR 29-NOV-2001; 2001US-0354370P.
PR 12-APR-2002; 2002US-0372246P.
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX Aziz N, Murray R;
PI WPI; 2003-093161/08.
DR N-PSDB; ABX76182.
DR
XX Detecting a lung cancer-associated transcript in a cell from a patient
PT for treating lung cancer, by contacting a biological sample from the
PT patient with a polynucleotide that exhibits increased or decreased
XX expression in lung cancer.
XX
PS Claim 27; Page 232; 453pp; English.
XX
XX The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hyperreactivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the
CC invention
CC
SQ Sequence 746 AA;
XX
XX Query Match 99.8%; Score 3802; DB 6; Length 746;
XX Best Local Similarity 99.9%; Pred. No. 2e-281;
XX Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



Db 301 ATPNTYKRNKTEALDNKPCGPOCYOHLGAKPAALATBAIRIKTPPKRRGRRLPN 360  
 Qy 361 NSSRSTPTINTLESKOTDSDREAGTGTGENDKXEEKKDETSSSSSEANRCOTPIKM 420  
 Db 361 NSSRSTPTINTLESKOTDSDREAGTGTGENDKXEEKKDETSSSSSEANRCOTPIKM 420  
 Qy 421 KPNIPPEVWESGABASMFRLVIGTYDNFCAIARLIGTKCQVYEFVRESIIAPA 480  
 Db 421 KPNIPPEVWESGABASMFRLVIGTYDNFCAIARLIGTKCQVYEFVRESIIAPA 480  
 Qy 481 PAEDVDTPPKRRKRLMAHCRKIQLKDGSSNHYNYOPCDHPROPDSSCPCTVIAQ 540  
 Db 481 PAEDVDTPPKRRKRLMAHCRKIQLKDGSSNHYNYOPCDHPROPDSSCPCTVIAQ 540  
 Qy 541 NFCEKFCQCSSECCNRPFRCRCCKAQCNTKQCPCTLAVERCDPDLCTLGAADHDSKXVS 600  
 Db 541 NFCEKFCQCSSECCNRPFRCRCCKAQCNTKQCPCTLAVERCDPDLCTLGAADHDSKXVS 600  
 Qy 601 CKNGSIOGSKKHLAPSDVAGMGIPTKDPVQKNBFISEYCGRIISODEADRGRKYDK 660  
 Db 601 CKNGSIOGSKKHLAPSDVAGMGIPTKDPVQKNBFISEYCGRIISODEADRGRKYDK 660  
 Qy 661 YMCSEFLPNLNDPVVDATRKGNKIRPANHSVNPCTAKV 699  
 Db 661 YMCSEFLPNLNDPVVDATRKGNKIRPANHSVNPCTAKV 699  
 RESULT 3  
 ADG10642  
 ID ADG10642 standard; protein; 746 AA.  
 AC ADG10642;  
 XX 26-FEB-2004 (first entry)  
 DT Human STAT6-activating protein, SEQ ID NO:232.  
 DE Human STAT6-activating protein, SEQ ID NO:232.  
 XX  
 KW Human; STAT6-activating protein; drug screening; activator; inhibitor;  
 KW allergic disease; inflammation; autoimmune disease; diabetes;  
 KW hyperlipidaemia; cancer; infection; HIV infection;  
 KW human immunodeficiency; cancer; Th1 hyperfunction; antiallergic;  
 KW antiinflammatory; antidiabetic; antihypertensive; antileukemic; anti-HIV;  
 KW cyostatic; gene therapy; antisense therapy; ribozyme therapy.  
 OS Homo sapiens.  
 XX  
 PN WO200296943-A1.  
 PV 05-DEC-2002.  
 PD 22-MAY-2002; 2002WO-JP004949.  
 PF 25-MAY-2001; 2001JP-00157043.  
 PR 30-AUG-2001; 2001JP-00260681.  
 PR 10-OCT-2001; 2001JP-00313175.  
 XX  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 XX  
 PI Honda G, Matcueda A, Muramatsu S, Ishizawa K;  
 DR N-PSDB; ADG10641.  
 DR N-PSDB; ADG10641.  
 XX  
 PT STAT6-activating proteins and encoded genes, applicable in diagnosis of  
 PT and developing drugs to treat allergic diseases, inflammations,  
 PT autoimmune diseases, diabetes, hyperlipidaemia, infections e.g. HIV, and  
 PT cancer.  
 XX  
 PS Claim 1; SEQ ID NO 232; 2080bp; Japanese.  
 XX  
 CC The invention relates to 242 human STAT6-activating proteins and cDNAs  
 CC encoding them (ADG10411-ADG10894) and to sequences with 95% or more  
 CC homology to the STAT6-activating proteins and their encoding nucleic

CC acids. The invention also relates to recombinant vectors and host cells  
 CC comprising a STAT6-activating protein-encoding nucleic acid; the  
 CC recombinant production of a STAT6-activating protein; an antibody  
 CC specific for a STAT6-activating protein; antisense oligonucleotides and  
 CC ribozymes targeted to nucleic acids encoding a STAT6-activating protein;  
 CC methods of screening for activators or inhibitors of STAT6-activating  
 CC proteins; drug compositions comprising a modulator of STAT6-activating  
 CC protein activity or expression; and methods of treating patients by  
 CC administration of the drug compositions. The STAT6-activating proteins,  
 CC nucleic acids encoding them, and modulators of their activity or  
 CC expression are useful in the diagnosis and treatment of allergic  
 CC diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia,  
 CC cancer, infections (e.g., HIV), cancer and disorders associated with Th1  
 CC hyperfunction. The present sequence is related to the invention.  
 XX  
 SQ Sequence 746 AA;  
 Query Match 99.8%; Score 3802; DB 7; Length 746;  
 Best Local Similarity 99.9%; Pred. No. 2e-281;  
 Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 MGOTGKSEKGPVCMRRKRVKSEYMLRQLKFRPADVKSMSNNRQKILERTBIANGW 60  
 Db 1 MGOTGKSEKGPVCMRRKRVKSEYMLRQLKFRPADVKSMSNNRQKILERTBIANGW 60  
 Qy 61 KORRIOPHILTSVSLRGTRGCVTSGLDPPVQVPLKTLNAAVPMVMSPLQONF 120  
 Db 61 KORRIOPHILTSVSLRGTRGCVTSGLDPPVQVPLKTLNAAVPMVMSPLQONF 120  
 Qy 121 MVEDETVLANIPYMGDEVLDQGTFFIELINQYGVKYGDRGCFINDEIVELVNALGQ 180  
 Db 121 MVEDETVLANIPYMGDEVLDQGTFFIELINQYGVKYGDRGCFINDEIVELVNALGQ 180  
 Qy 121 MVEDETVLANIPYMGDEVLDQGTFFIELINQYGVKYGDRGCFINDEIVELVNALGQ 180  
 Db 121 MVEDETVLANIPYMGDEVLDQGTFFIELINQYGVKYGDRGCFINDEIVELVNALGQ 180  
 Qy 181 YNDDDDDDGDDPBEREKQDLHDHDKESRPPKRPSPDKIFEAISMPDPKGTAEEL 240  
 Db 181 YNDDDDDDGDDPBEREKQDLHDHDKESRPPKRPSPDKIFEAISMPDPKGTAEEL 240  
 Qy 241 KEKKEITLQOQLPQALPPECPTNIDGNPAKSVQEGOSHSHTPLFCRCRCFYDCTLHFFH 300  
 Db 241 KEKKEITLQOQLPQALPPECPTNIDGNPAKSVQEGOSHSHTPLFCRCRCFYDCTLHFFH 300  
 Qy 301 ATPNTYKRNKTEALDNKPCGPOCYOHLGAKPAALATBAIRIKTPPKRRGRRLPN 360  
 Db 301 ATPNTYKRNKTEALDNKPCGPOCYOHLGAKPAALATBAIRIKTPPKRRGRRLPN 360  
 Qy 361 NSSRSTPTINTLESKOTDSDREAGTGTGENDKXEEKKDETSSSSSEANRCOTPIKM 420  
 Db 361 NSSRSTPTINTLESKOTDSDREAGTGTGENDKXEEKKDETSSSSSEANRCOTPIKM 420  
 Qy 421 KPNIPPEVWESGABASMFRLVIGTYDNFCAIARLIGTKCQVYEFVRESIIAPA 480  
 Db 421 KPNIPPEVWESGABASMFRLVIGTYDNFCAIARLIGTKCQVYEFVRESIIAPA 480  
 Qy 481 PAEDVDTPPKRRKRLMAHCRKIQLKDGSSNHYNYOPCDHPROPDSSCPCTVIAQ 540  
 Db 481 PAEDVDTPPKRRKRLMAHCRKIQLKDGSSNHYNYOPCDHPROPDSSCPCTVIAQ 540  
 Qy 541 NFCEKFCQCSSECCNRPFRCRCCKAQCNTKQCPCTLAVERCDPDLCTLGAADHDSKXVS 600  
 Db 541 NFCEKFCQCSSECCNRPFRCRCCKAQCNTKQCPCTLAVERCDPDLCTLGAADHDSKXVS 600  
 Qy 601 CKNGSIOGSKKHLAPSDVAGMGIPTKDPVQKNBFISEYCGRIISODEADRGRKYDK 660  
 Db 601 CKNGSIOGSKKHLAPSDVAGMGIPTKDPVQKNBFISEYCGRIISODEADRGRKYDK 660  
 Qy 661 YMCSEFLPNLNDPVVDATRKGNKIRPANHSVNPCTAKV 699  
 Db 661 YMCSEFLPNLNDPVVDATRKGNKIRPANHSVNPCTAKV 699  
 RESULT 4  
 ADN39392  
 ID ADN39392 standard; protein; 746 AA.

XX ADN39392;  
AC  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: B76.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;  
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KW inflammatory disease; autoimmune disease;  
KW retinal neovascularization syndrome; scarring; uterine fibroid;  
KW detection; diagnosis; prognosis; drug screening; drug targeting;  
KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
KW vlnutary; gene therapy; vaccine.  
XX  
XX Homo sapiens.  
XX  
XX WO2003042661-A2.  
XX  
XX  
XX 22-MAY-2003.  
XX  
XX  
XX 13-NOV-2002; 2002WO-US036610.  
XX  
XX  
XX 13-NOV-2001; 2001US-0350666P.  
XX 21-NOV-2001; 2001US-0332464P.  
XX 29-NOV-2001; 2001US-0334393P.  
XX 03-DEC-2001; 2001US-035394P.  
XX 14-DEC-2001; 2001US-0340376P.  
XX 08-JAN-2002; 2002US-0347231P.  
XX 10-JAN-2002; 2002US-0347349P.  
XX 08-FEB-2002; 2002US-0355250P.  
XX 13-FEB-2002; 2002US-0356714P.  
XX 20-FEB-2002; 2002US-0359077P.  
XX 29-MAR-2002; 2002US-036809P.  
XX 04-APR-2002; 2002US-0370110P.  
XX 12-APR-2002; 2002US-0372246P.  
XX 05-JUN-2002; 2002US-036614P.  
XX 16-JUL-2002; 2002US-036839P.  
XX 22-JUL-2002; 2002US-0397775P.  
XX 22-JUL-2002; 2002US-0397845P.  
XX 09-SEP-2002; 2002US-0409450P.  
XX  
XX (BOSB-) BOS BIOTECHNOLOGY INC.  
XX  
XX Afez D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Heyezi PA;  
PI Mack DH, Murray R, Watson SR, Wilson KB, Zlotnik A;  
XX WPI; 2003-468649/44.  
XX N-PSDB; ADN39391.  
XX  
XX  
XX Determining the presence or absence of a pathological cell in a patient,  
XX useful for diagnosing, prognosing or treating cancer, comprises detecting  
XX a nucleic acid in a biological sample.  
XX  
XX  
XX Claim 12; SEQ ID NO B76; 1385pp; English.  
XX  
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
XX whose expression is upregulated or downregulated in specific cancers or  
XX other diseases such as angiogenic or fibrotic disorders, and to methods  
XX of determining the presence or absence of a pathological cell in a  
XX patient by detecting a nucleic acid at least 80% identical to those of  
XX the invention or by detecting a polypeptide of the invention. The  
XX invention also relates to expression vectors and host cells comprising a  
XX nucleic acid of the invention; antibodies which specifically bind a  
XX polypeptide of the invention; use of such antibodies for drug targeting;  
XX and methods of screening for modulators of activity or expression of the  
XX polypeptides and nucleic acids. The nucleic acids, polypeptides,  
XX antibodies and methods are useful for diagnosing, prognosing and treating  
XX cancer and other conditions such as psoriasis, ischaemia, heart disease,  
XX atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
XX neovascularization syndromes, scarring and uterine fibroids. They may  
XX also be useful in wound healing and in contraception. The present  
XX sequence represents a polypeptide of the invention.

XX  
SQ Sequence 746 AA;  
Query Match 99.8%; Score 3802; DB 7; Length 746;  
Best Local Similarity 99.9%; Pred. No. 2e-281;  
Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 1 MGQTGKSKSEKPYCWRKRVKSEYMLRQLKFRFRADEVKSMFSNNOKILERTYLLNQEW 60  
1 MGQTGKSKSEKPYCWRKRVKSEYMLRQLKFRFRADEVKSMFSNNOKILERTYLLNQEW 60  
DB  
QY 61 KQRIQPVHLLTVSSLRGRCSVTSDDLFPQOVPLKTLNVAASPIVMYSPIQONF 120  
61 KQRIQPVHLLTVSSLRGRCSVTSDDLFPQOVPLKTLNVAASPIVMYSPIQONF 120  
DB  
QY 121 MVEDEVLANIPYMGDEVLDQDGTFLBELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
121 MVEDEVLANIPYMGDEVLDQDGTFLBELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
DB  
QY 121 MVEDEVLANIPYMGDEVLDQDGTFLBELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
121 MVEDEVLANIPYMGDEVLDQDGTFLBELIKNYDGKVGDRGCFINDEIFVELVNALGQ 180  
DB  
QY 181 YNDDDDDDGDDPEBBREKQDLIEDHRDDKSRPPPKFPBDKILKAISSMFPDKTAEEL 240  
181 YNDDDDDDGDDPEBBREKQDLIEDHRDDKSRPPPKFPBDKILKAISSMFPDKTAEEL 240  
DB  
QY 241 KERYKELTEQQLFGALPPECPTNIDGPNKSVGRGSLSHFHTLFCRCFKYDCPLHPFH 300  
241 KERYKELTEQQLFGALPPECPTNIDGPNKSVGRGSLSHFHTLFCRCFKYDCPLHPFH 300  
DB  
QY 301 ATPYTYRKNTETALDNKPCGPOCYOHLGAKFPAALTAERLKTTPKRPGRGRRLPN 360  
301 ATPYTYRKNTETALDNKPCGPOCYOHLGAKFPAALTAERLKTTPKRPGRGRRLPN 360  
DB  
QY 361 NSSRPSTPTINVLBSKQTDSDREAGTGTGENDDKBEEKDGTSSSSSEANSKQTPYK 420  
361 NSSRPSTPTINVLBSKQTDSDREAGTGTGENDDKBEEKDGTSSSSSEANSKQTPYK 420  
DB  
QY 421 KPNIEPENVENWGAESMFRVILGYTNDPFCALRIIGTKROYVEFVKSSTIAPA 480  
421 KPNIEPENVENWGAESMFRVILGYTNDPFCALRIIGTKROYVEFVKSSTIAPA 480  
DB  
QY 481 PAEDVDTPPKKRRKRLMAHCRKIQLKKGSSNMHYVYQPCDHPQPCDSSCPVIAQ 540  
481 PAEDVDTPPKKRRKRLMAHCRKIQLKKGSSNMHYVYQPCDHPQPCDSSCPVIAQ 540  
DB  
QY 541 NFCEKFCQSSSECONFPGCRCAQCNTKQPCYLAVRECDPILCLTGADADWDSDKNVS 600  
541 NFCEKFCQSSSECONFPGCRCAQCNTKQPCYLAVRECDPILCLTGADADWDSDKNVS 600  
DB  
QY 601 CRKCSIORGSKRHLLAPSDVAGMGTIKDPVQKNEFISYCGEIIISQDEADRGKYYDK 660  
601 CRKCSIORGSKRHLLAPSDVAGMGTIKDPVQKNEFISYCGEIIISQDEADRGKYYDK 660  
DB  
QY 661 YMCSPFLNANDFVVDATRKANKIRFANHSVNPYCAKV 699  
661 YMCSPFLNANDFVVDATRKANKIRFANHSVNPYCAKV 699  
DB  
QY 661 YMCSPFLNANDFVVDATRKANKIRFANHSVNPYCAKV 699  
661 YMCSPFLNANDFVVDATRKANKIRFANHSVNPYCAKV 699  
DB  
RESULT 5  
ADP95920  
ID ADP95920 standard; protein; 751 AA.  
XX  
AC ADP95920;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DB Human E2H2 (enhancer of zeste homologue 2), SEQ ID NO:1.  
XX  
XX Human; E2H2; enhancer of zeste homologue 2; chromosome 7q35;  
KW histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
KW inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
KW lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
KW prophylaxis; prevention; diagnosis; cytostatic; enzyme.  
XX  
XX Homo sapiens.

XX MO2004058969-A1.  
 PN 15-JUL-2004.  
 PD 22-DEC-2003; 2003WO-JP016417.  
 XX 24-DEC-2002; 2002JP-00373144.  
 XX (TAKEDA CHEM IND LTD.  
 PA HIKICHI Y, NISHIZAWA S;  
 PI WPI: 2004-525885/50.  
 DR N-PEDB; ADP95921, ADP95922.  
 XX  
 PT Novel prophylactic and therapeutic agent of cancer, having compound or  
 its salt that inhibits activity of enhancer of zeste Drosophila homolog2  
 protein, useful for treating cancer such as lung cancer, breast cancer.  
 XX  
 PS Claim 1; SEQ ID NO 1; 102pp; Japanese.

CC The invention relates to an agent for the prophylaxis and therapy of  
 CC cancer comprising a compound which inhibits the activity or expression of  
 CC human EZH2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
 CC EZH2 inhibitors include anti-EZH2 antibodies and EZH2 antisense nucleic  
 CC acids. The invention also relates to a kit for screening for inhibitors  
 CC of EZH2 activity or expression. EZH2 is a component of a complex with  
 CC histone methyltransferase activity which methylates histone H3 on lysine  
 CC residues at position 9 and/or 27. It has been found to be deregulated in  
 CC various tumours, and acts as a mediator of tumorigenesis, being able to  
 CC reverse repression of cyclin A which in turn mediates cell cycle  
 CC advancement. The agent of the invention can be used for preventing,  
 CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
 CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
 CC in drug screening. The agent of the invention is safe and less toxic than  
 CC prior art cancer prophylactic/therapeutic agents. The present sequence  
 CC represents human EZH2, which is referred to in the claims.

XX Sequence 751 AA;

Query Match 99.5%; Score 3787.5; DB 8; Length 751;  
 Best Local Similarity 99.1%; Pred. No. 2.6e-280;  
 Matches 698; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 MGQTGKSEKGPVCKRKKVSEYMLRQLKRRADRYKSMFSSNRQKILERTILNGEW 60  
 DB 1 MGQTGKSEKGPVCKRKKVSEYMLRQLKRRADRYKSMFSSNRQKILERTILNGEW 60  
 QY 61 KORRIQPHILITSSVSLAGTRCSTSLDLPPTQVITPKLTANAVAPIMTSGSPLOQNF 120  
 DB 61 KORRIQPHILITSSVSLAGTRCSTSLDLPPTQVITPKLTANAVAPIMTSGSPLOQNF 120  
 QY 121 MVEDETVLANIPYMGDEVLDODGFTIEELIKNYGKVGDRRCGINDIEIVELVNALGQ 180  
 DB 121 MVEDETVLANIPYMGDEVLDODGFTIEELIKNYGKVGDRRCGINDIEIVELVNALGQ 180  
 QY 181 YNDDDDDDGDDPSEERBEKQKOLBDHRDKESRPPRPKPSDKITFEALISMFPPDKGTAEL 240  
 DB 181 YNDDDDDDGDDPSEERBEKQKOLBDHRDKESRPPRPKPSDKITFEALISMFPPDKGTAEL 240  
 QY 241 KEKTKELTEQOLPGALPPECTPNIDGPNVAKSVOREQSASHFTLFCRCFCYKDCFLHKKC 300  
 DB 241 KEKTKELTEQOLPGALPPECTPNIDGPNVAKSVOREQSASHFTLFCRCFCYKDCFLHKKC 300  
 QY 298 --PFIATNTYKRNKTEFALNKPQGCYQHLISGAKSEFAALTAERTKTPPKRPGGRR 355  
 DB 301 NYSHATNTYKRNKTEFALNKPQGCYQHLISGAKSEFAALTAERTKTPPKRPGGRR 360  
 QY 356 GRLPNNSRSPSTPTINVLSEKDTSDREAGTETGGENNDKEEKKOETSSSSSEANSRCQ 415  
 DB 361 GRLPNNSRSPSTPTINVLSEKDTSDREAGTETGGENNDKEEKKOETSSSSSEANSRCQ 420

QY 416 TPIMKKPIBPENVENSGAEMRVLIGTYDNPCALARTLIGTKRGVYBFRVKSS 475  
 DB 421 TPIMKKPIBPENVENSGAEMRVLIGTYDNPCALARTLIGTKRGVYBFRVKSS 480  
 QY 476 IIAAPADVDVTPPRKKRKRRLWAAMCRKIQLKKDGSNNHYVYQPCDHRQPCDSSCP 535  
 DB 481 IIAAPADVDVTPPRKKRKRRLWAAMCRKIQLKKDGSNNHYVYQPCDHRQPCDSSCP 540  
 QY 536 CVIAQNPCEKCCQSSSEQRNFPFCRCCKAQCNTQCPCTYLAVRCDPDLCTCGAADWD 595  
 DB 541 CVIAQNPCEKCCQSSSEQRNFPFCRCCKAQCNTQCPCTYLAVRCDPDLCTCGAADWD 600  
 QY 596 SKVNSCKNCSIQSGSKHLLAPSDVAGWGIFFIDPQKNEFISEYCGEIIISQDEADRG 655  
 DB 601 SKVNSCKNCSIQSGSKHLLAPSDVAGWGIFFIDPQKNEFISEYCGEIIISQDEADRG 660  
 QY 656 KYVDKYMCSFLFNINDFVVDATKGNKIRPANSVNPNCYAKV 699  
 DB 661 KYVDKYMCSFLFNINDFVVDATKGNKIRPANSVNPNCYAKV 704

#### RESULT 6

ID ADB65008 standard; protein; 707 AA.  
 AC ADB65008;  
 DT 04-DEC-2003 (first entry)  
 DE Human protein encoded by clone PUAEN20003120.  
 XX Human, pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.  
 OS Homo sapiens.  
 XX EP1308459-A2.  
 PN 07-MAY-2003.  
 PD 28-MAR-2002; 2002EP-00007401.  
 XX 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 DR WPI: 2003-450961/43.  
 DR N-PEDB; ADB65038.  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 marker or medicines for regulation of their expression and activity, or  
 as targets of gene therapy.  
 XX Claim 1; Page: 222pp; English.  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or peptide  
 CC of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transfectant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe

CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a protein of the invention. Note: Some of the  
CC sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX  
XX  
SQ Sequence 707 AA:

Query Match 93.7%; Score 3569.5; DB 7; Length 707;  
Best Local Similarity 94.3%; Pred. No. 1.1e-263;  
Matches 659; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

QY 1 MGOTGKSKGKPCVCKRKRKVSRYRLRLKFRADRYKSWFSNNRQKILERTIILNQEW 60  
DB 1 MGOTGKSKGKPCVCKRKRKVSRYRLRLKFRADRYKSWFSNNRQKILERTIILNQEW 60  
QY 61 KORRIQPVHILTSVSLRGTRSCVTSDDLDPFTQVILPLKTLNAVASVPIMYSNPLQONF 120  
DB 61 KORRIQPVHILTSVSLRGTRSCVTSDDLDPFTQVILPLKTLNAVASVPIMYSNPLQONF 120  
QY 121 MVBETVLHNPVWGDVTLQDDGTFFIELIKNYGKYGHGREGCFINDEIFVELVNALGQ 180  
DB 83 -VEDETVALHNPVWGDVTLQDDGTFFIELIKNYGKYGHGREGCFINDEIFVELVNALGQ 141  
QY 181 YNDDDDDDGDDPEREREKQDLERHDDKESRPPKPSDKITREALISMFPPDKGTAEEL 240  
DB 142 YNDDDDDDGDDPEREREKQDLERHDDKESRPPKPSDKITREALISMFPPDKGTAEEL 201  
QY 241 KEKTKELTEQQLPGALPPECTPNTIDGNPAKSVQREOSLHSPHTLFCRCFKYDCFLHPFH 300  
DB 202 KEKTKELTEQQLPGALPPECTPNTIDGNPAKSVQREOSLHSPHTLFCRCFKYDCFLHPFH 261  
QY 301 ATPPTTKRKTETLADNKPCCGQCYOHLGAKERPAALTAARITTPPRRPGRRGRRLPN 360  
DB 262 ATPPTTKRKTETLADNKPCCGQCYOHLGAKERPAALTAARITTPPRRPGRRGRRLPN 321  
QY 361 NSSRPTPTINVLKSDTDSRBAETGTGENNDKEBEKKDRTSSSEANSRCQTPYIKM 420  
DB 322 NSSRPTPTINVLKSDTDSRBAETGTGENNDKEBEKKDRTSSSEANSRCQTPYIKM 381  
QY 421 KPNIEPPENVWMSGABASMPFVLLIGTYYNFCALIRLLIGTYTCQVYEFVYKSSIIAPA 480  
DB 382 KPNIEPPENVWMSGABASMPFVLLIGTYYNFCALIRLLIGTYTCQVYEFVYKSSIIAPA 441  
QY 481 PABVNDTPPRKKRKHRLMAHCKRIOLKQDSSNHVYVNPCCHPPOCCSSPCVYIAQ 540  
DB 442 PABVNDTPPRKKRKHRLMAHCKRIOLKQDSSNHVYVNPCCHPPOCCSSPCVYIAQ 501  
QY 541 NFCEKFCQCCSSSECONRPFQGCCKAQACNTKQCCPTLAVERCDPDLCLTCGADHDSKNVS 600  
DB 502 NFCEKFCQCCSSSECONRPFQGCCKAQACNTKQCCPTLAVERCDPDLCLTCGADHDSKNVS 561  
QY 601 CKNCISIGSGKXKLLAPSDVANGIFIKDPVQKNEPISYCGIISODEADRGRKYDK 660  
DB 562 CKNCISIGSGKXKLLAPSDVANGIFIKDPVQKNEPISYCGIISODEADRGRKYDK 621  
QY 661 YMGSPFLNANDPVYDATRKGKRIKPNASHVNPNCYAKV 699  
DB 622 YMGSPFLNANDPVYDATRKGKRIKPNASHVNPNCYAKV 660

RESULT 7  
ADG10764  
ID ADG10764 standard; protein: 517 AA.XX  
AC ADG10764;  
XX  
DT 26-FEB-2004 (first entry)

DE Human STAT6-activating protein, SEQ ID NO:354.

KW Human; STAT6-activating protein; drug screening; activator; inhibitor;  
KW allergic diseases; inflammation; autoimmune disease; diabetes;  
KW hyperlipidaemia; cancer; infection; HIV infection;  
KW human immunodeficiency; cancer; Th1 hyperfunction; antiallergic;  
KW antiinflammatory; antidiabetic; antilipemic; antiinfective; anti-HIV;  
KW cytostatic; gene therapy; antisense therapy; ribozyme therapy.XX  
OS Homo sapiens.XX  
PN WO200296943-A1.XX  
PD 05-DEC-2002.XX  
PF 22-MAY-2002; 2002WO-JP004949.XX  
PR 25-MAY-2001; 2001JP-00157043.

PR 30-AUG-2001; 2001JP-00260681.

PR 10-OCT-2001; 2001JP-00313175.

XX  
PA (ASAH ) ASAH KASEI KOGYO KK.XX  
PI Honda G, Matsuda A, Muramatsu S, Ishizawa K;XX  
PI WPI, 2003-140442/13.XX  
DR N-PSDB; ADG10763.XX  
PT STAT6-activating proteins and encoded genes, applicable in diagnosis of

PT and developing drugs to treat allergic diseases, inflammations,

PT autoimmune diseases, diabetes, hyperlipidemia, infections e.g. HIV, and

PT cancer.

XX  
PS Claim 1; SEQ ID NO 354; 2080bp; Japanese.XX  
PS The invention relates to 242 human STAT6-activating proteins and cDNAsXX  
CC encoding them (ADG10411-ADG10894) and to sequences with 95% or moreXX  
CC homology to the STAT6-activating proteins and their encoding nucleicXX  
CC acids. The invention also relates to recombinant vectors and host cellsXX  
CC comprising a STAT6-activating protein-encoding nucleic acid; theXX  
CC recombinant production of a STAT6-activating protein; an antibodyXX  
CC specific for a STAT6-activating protein; antisense oligonucleotides andXX  
CC ribozymes targeted to nucleic acids encoding a STAT6-activating protein;XX  
CC methods of screening for activators or inhibitors of STAT6-activatingXX  
CC proteins; drug compositions comprising a modulator of STAT6-activatingXX  
CC protein activity or expression; and methods of treating patients byXX  
CC administration of the drug compositions. The STAT6-activating proteins,XX  
CC nucleic acids encoding them, and modulators of their activity orXX  
CC expression are useful in the diagnosis and treatment of allergicXX  
CC diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia,XX  
CC cancer, infections (e.g., HIV), cancer and disorders associated with Th1XX  
CC hyperfunction. The present sequence is related to the invention.XX  
SQ Sequence 517 AA:

Query Match 68.3%; Score 2602; DB 7; Length 517;  
Best Local Similarity 100.0%; Pred. No. 6.3e-190; Indels 0; Gaps 0;  
Matches 470; Conservative 0; Mismatches 0;

QY 230 MPPDKGTAEELKEXYKELTEQQLPGALPPECTPNTIDGNPAKSVQREOSLHSPHTLFCRCRC 289  
DB 1 MPPDKGTAEELKEXYKELTEQQLPGALPPECTPNTIDGNPAKSVQREOSLHSPHTLFCRCRC 60  
QY 290 FKYDCFLPHAPNTYKRNKTETLADNKPCCGQCYOHLGAKERPAALTAARITKTPPKR 349  
DB 61 FKYDCFLPHAPNTYKRNKTETLADNKPCCGQCYOHLGAKERPAALTAARITKTPPKR 120  
QY 350 PGRRRGRRLPNNSRPTPTINVLKSDTDSRBAETGTGENNDKEBEKKDRTSSSSSE 409

DB 121 FGRRRGLPNNSSRPSTPTINVLSEKOTDSREAGTGTGNNNDKEBEKKDERTSSSE 180  
 QY 410 ANSRQGTITKKNPENTPEPPENWWSGABASMPFVLLIGTYDNFCALRLIGTTCQVYEF 469  
 DB 181 ANSRQGTITKKNPENTPEPPENWWSGABASMPFVLLIGTYDNFCALRLIGTTCQVYEF 240  
 QY 470 RVKSSIIAPAPAEVDTPPRKKRKHLMANCKIKLQKDGSSNNHYNTQPCDHPOR 529  
 DB 241 RVKSSIIAPAPAEVDTPPRKKRKHLMANCKIKLQKDGSSNNHYNTQPCDHPOR 300  
 QY 530 CDSGCPYIAONFCERFCQCSSECONRPPGCRCKAQCTKQPCYLAARECDPDLCTCG 589  
 DB 301 CDSGCPYIAONFCERFCQCSSECONRPPGCRCKAQCTKQPCYLAARECDPDLCTCG 360  
 QY 590 AADHWSKNVSCKNCSIORGSKHLLAPSDVAGNGITFKDPVQKNPFISEYCGEIIISQD 649  
 DB 361 AADHWSKNVSCKNCSIORGSKHLLAPSDVAGNGITFKDPVQKNPFISEYCGEIIISQD 420  
 QY 650 EADRRGKYVDKYMCSFLFNLNNDPVYDATRKGNKIRFANHSVNPICYAKV 699  
 DB 421 EADRRGKYVDKYMCSFLFNLNNDPVYDATRKGNKIRFANHSVNPICYAKV 470

RESULT 8  
 ADX05848  
 ID ADX05848 standard; protein; 747 AA.  
 AC ADX05848;  
 DT 21-APR-2005 (first entry)  
 DE Cyclin-dependent kinase modulation biomarker SEQ ID NO 413.  
 KM cytosolic; cyclin-dependent kinase; cdk; biomarker.  
 OS Homo sapiens.  
 PN WO2005012875-A2.  
 PD 10-FEB-2005.  
 PF 29-JUL-2004; 2004WO-US024424.  
 PR 29-JUL-2003; 2003US-0490890P.  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PI Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;  
 DR WPI; 2005-163068/17.  
 DR N-PSDB; ADX05847.  
 XX Biomarkers useful for predicting or determining the response of a mammal  
 PT to a cancer treatment comprising administration of a modulator of cyclin-  
 dependent kinase activity.  
 XX Claim 5; SEQ ID NO 413; 141pp; English.

This invention describes a novel method of predicting or determining whether a mammal will respond or is responding to an anti-cancer agent that modulates cyclin-dependent kinase (cdk) activity. The method comprises measuring the level of one or more biomarkers selected from 2774 biomarkers given in the specification (nucleotide sequence SEQ ID NO:1246 (Genbank BSR W28729) is especially preferred). The method of the invention is utilized in a kit for determining or predicting whether a patient would be susceptible or resistant to treatment by an agent modulating cdk activity. The invention also describes a method for utilizing individualized genetic profiles for treating diseases and disorders based on patient's response and molecular level, specialized microarrays comprising the biomarkers described, antibodies directed against the biomarkers and a cell culture model to identify biomarkers. The cdk modulator is preferably N-5-[5-(1,1-dimethylethyl)-2-

CC oxazoly[1]methyl[thio]-2-thiazoly[1-4-piperidine carboxamide, 0.5-L-  
 CC tartaric acid salt. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published.pct\_sequences. This  
 CC sequence represents a biomarker used in the method of the invention.  
 XX  
 SQ Sequence 747 AA;  
 Query Match 62.6%; Score 2384.5; DB 9; Length 747;  
 Best Local Similarity 63.1%; Pred. No. 4,3e-173;  
 Matches 449; Conservative 89; Mismatches 121; Indels 53; Gaps 10;  
 QY 15 WRKVKSEYVRLRLQKPRADBYKSMPSNRKILRRTILNQBKORRLOPHILTSV 74  
 DB 15 WRKVKSEYVRLRLQKPRADBYKSMPSNRKILRRTILNQBKORRLOPHILTSV 74  
 QY 75 SSLGTRBCSVTSDL-DEPTQVILPLKTLMAVASVPIWMSPLQONFVDEYLAHNPY 133  
 DB 75 SGHPFLAKCTIBSIFPGPASQMLMRSLNTVALVPIWMSPLQONFVDEYLAHNPY 134  
 QY 134 MGDEVLDQDGTIFIELIKNTDGKYGDR-----CGFINDEIFVELVNAAGYND----- 185  
 DB 135 MGDEVKEDDTFIELINNTDGKYGHEMTPGSVLISDAVFLVYALANQYSDBERGH 194  
 QY 186 -----DDDDGDPPERBEREKOKLDHNDKESRPPKPSDKIFPAISMPFDKGT 237  
 DB 195 NDTSDGKODSKEDLPYTRKKRKHLMANCKIKLQKDGSSNNHYNTQPCDHPOR 250  
 QY 238 BELKRYKYLTEQOLPGALPEECTPNIDGPAKSVQREOSLHSPHTLFCRACFYDCFLH 297  
 DB 251 DDMKRYRBLTMSDPNALPQCTPNIDGPAKSVQREOSLHSPHTLFCRACFYDCFLH 310  
 QY 298 PFATPTTYTKRKTETALDNKPCGPOCYOHLKAKERPAALTAIRITPPRPGRRGR 357  
 DB 311 PFATPTTYTKRKTETALDNKPCGPOCYOHLKAKERPAALTAIRITPPRPGRRGR 364  
 QY 358 LPNNSRPSTPTINVLSEKOTDSREAGTGTGNNNDKEBEKKDERTSSSEANRSQCT 416  
 DB 365 HHIVSASCNSASAVAKETREGSDSDRTG-----NDMASSSEANRSQCT 409  
 QY 417 PIRKKPIBPP-----ENWWSGABASMPFVLLIGTYDNFCALRLIGTTCQVY 467  
 DB 410 PTKQKASPAPQCLVBAPEBPVEMTGABESLIFPVHGTGFNNNCIARLLGTTCQVY 469  
 QY 468 EFRVYSSIIAPAPAEVDTPPRKKRKHLMANCKIKLQKDGSSNNHYNTQPCDHPOR 527  
 DB 470 QFAVKE-SLILKLPTEDELMTPSQKKRKHLMANCKIKLQKDGSSNNHYNTQPCDHPOR 528  
 QY 528 QPCDSSGCPYIAONFCERFCQCSSECONRPPGCRCKAQCTKQPCYLAARECDPDLCT 587  
 DB 529 RPCDSTPCIMTQNFCEKFCQCSSECONRPPGCRCKAQCTKQPCYLAARECDPDLCT 588  
 QY 588 CGAADHWSKNVSCKNCSIORGSKHLLAPSDVAGNGITFKDPVQKNPFISEYCGEIIIS 647  
 DB 589 CGASEHNDCKVSCKNCSIORGSKHLLAPSDVAGNGITFKDSVQKNPFISEYCGEIIIS 648  
 QY 648 QDEADRRGKYVDKYMCSFLFNLNNDPVYDATRKGNKIRFANHSVNPICYAKV 699  
 DB 649 QDEADRRGKYVDKYMCSFLFNLNNDPVYDATRKGNKIRFANHSVNPICYAKV 700

RESULT 9  
 ABU11057  
 ID ABU11057 standard; protein; 723 AA.  
 AC ABU11057;  
 DT 05-FEB-2003 (first entry)  
 DE Human protein NOV21.  
 KM Human; NOVX; adrenoleukodystrophy; haemophilia; stroke; VHL; congenital adrenal hyperplasia; haemophilia; hypercoagulation;

KW idiopathic thrombocytopenic purpura; autoimmune disease; allergy;  
KW immunodeficiencies; transplantation; Von Hippel-Lindau syndrome;  
KW Alzheimer's disease; tubular sclerosis; Parkinson's disease; epilepsy;  
KW Huntington's disease; cerebral palsy; Lesch-Nyhan syndrome; pain;  
KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety;  
KW behavioral disorder; addiction; neuroprotection; diabetes; AIDS;  
KW renal artery stenosis; interstitial nephritis; glomerulonephritis;  
KW polycystic kidney disease; systemic lupus erythematosus; IGA;  
KW renal tubular acidosis; immunoglobulin A nephropathy; hypercalcaemia;  
KW cirrhosis; transplantation; asthma; emphysema; scleroderma; GVHD;  
KW adult respiratory distress syndrome; graft versus host disease;  
KW lymphedema; fertility; pancreatitis; obesity; haemophilia; ulcer;  
KW anaemia; cancer; trauma; regeneration; infection.  
XX Homo sapiens.  
XX NC0200281629-A2.  
XX 17-OCT-2002.  
XX 03-APR-2002; 2002WO-US010522.  
XX 03-APR-2001; 2001US-0281086P.  
XX 03-APR-2001; 2001US-0281136P.  
XX 05-APR-2001; 2001US-0281863P.  
XX 05-APR-2001; 2001US-0281906P.  
XX 06-APR-2001; 2001US-0282020P.  
XX 10-APR-2001; 2001US-0282934P.  
XX 12-APR-2001; 2001US-0283512P.  
XX 19-APR-2001; 2001US-0285325P.  
XX 23-APR-2001; 2001US-0285890P.  
XX 24-APR-2001; 2001US-0286068P.  
XX 25-APR-2001; 2001US-0286292P.  
XX 27-APR-2001; 2001US-0287213P.  
XX 02-MAY-2001; 2001US-0288257P.  
XX 12-MAY-2001; 2001US-0281134P.  
XX 17-MAY-2001; 2001US-0291725P.  
XX 31-MAY-2001; 2001US-0294771P.  
XX 08-JUN-2001; 2001US-0296965P.  
XX 18-JUN-2001; 2001US-0299128P.  
XX 12-JUN-2001; 2001US-0305063P.  
XX 14-NOV-2001; 2001US-0332780P.  
XX 04-JAN-2002; 2002US-0345221P.  
XX 02-APR-2002; 2002US-00345221.  
XX (CURA-) CURAGEN CORP.  
XX Sphytek KA, Li L, Edinger SR, Ellerman K, Stone DJ, Malysankar UM,  
PI Shymkets RA, Guo X, Anderson DW, Patursajan M, Berghs C, Gerlach V,  
PI Taupier RJ, Pena CRA, Padigaru M, Liu Y, Burgess CE, Miller CE,  
PI Guevra VV, Kekuda R, Gorman L, Zernusen BD, Baumgartner JC,  
PI Tchernev VT, Vernet CM, Smithson G, Heyes MP, Shenoy SG, Liu X,  
PI Gangoli EA;  
XX WPI; 2003-046863/04.  
XX N-PSDB; ABX17535.  
XX New polypeptides, designated NOVX polypeptides, useful for treating  
PT hemophilia, idiopathic thrombocytopenic purpura, autoimmune disease,  
PT allergies, transplantation, Alzheimer's disease and stroke.  
XX Claim 1; Page 152; 320pp; English.  
XX The invention relates to an isolated NOVX polypeptide selected from NOV1-  
CC NOV27 polypeptides, a mature form of NOVX, a variant of NOVX or a  
CC fragment of NOVX. Also included are determining the presence or amount of  
CC NOVX in a sample (by using an antibody that immunospecifically bind to  
CC the polypeptide), determining the presence of or predisposition to  
CC disease associated with altered levels of NOVX in a first mammalian  
CC subject, identifying a potential therapeutic agent for use in the  
CC treatment of pathology related to aberrant expression of physiological  
CC interactions of NOVX, screening for a modulator of activity or of latency  
CC or predisposition to a pathology associated with NOVX, the nucleic acid

CC encoding NOVX, vectors and host cells. NOVX is useful for identifying an  
CC agent (a cellular receptor or downstream effector) that binds to NOVX.  
CC NOVX and NOVX nucleic acids are useful for treating or preventing NOVX-  
CC associated disorders in humans, and in the manufacture of a medicament  
CC for treating a NOVX related disease human disease e.g.  
CC adrenoleukodystrophy, congenital adrenal hyperplasia, haemophilia,  
CC hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune  
CC disease, allergies, immunodeficiencies, transplantation, Von Hippel-  
CC Lindau (VHL) syndrome, Alzheimer's disease, stroke, tubular sclerosis,  
CC Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy,  
CC Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,  
CC leukodystrophies, behavioural disorders, addiction, anxiety, pain,  
CC neuroprotection, diabetes, renal artery stenosis, systemic lupus  
CC glomerulonephritis, polycystic kidney disease, emphysema, IGA,  
CC erythematosus, renal tubular acidosis, immunoglobulin (Ig) A nephropathy,  
CC hypercalcaemia, cirrhosis, transplantation, asthma, emphysema,  
CC scleroderma, adult respiratory distress syndrome (ARDS), graft versus  
CC host disease (GVHD), lymphedema, fertility, pancreatitis, obesity,  
CC haemophilia, ulcers, anaemia, cancer, trauma, regeneration, and viral,  
CC bacterial or parasitic infections. The present sequence represents a NOVX  
CC protein  
XX  
XX Sequence 723 AA:  
Query Match 58.1%; Score 2212.5; DB 6; Length 723;  
Best Local Similarity 59.7%; Pred. No. 5,8e-160;  
Matches 425; Conservative 90; Mismatches 120; Indels 77; Gaps 11;  
QY 15 WRKRVSEVWRLQLRFRADDEVKSMFSSNRKILERTILNQEKORRLOPHILTSV 74  
15 WKRVSEVWRLQLRQLDANKAKLVANFPAKVEKQILNEBKAKRVGVGSMKV 74  
QY 75 SSLRGTRGCVTSDL-DPTQVLPKLTANAVSPIMYSWSPLOQFMVEDFTVLANIPY 133  
75 SGHPLKCTTBSIPGFAQSCHMLMSLTVALVPIYMSWSPLOQFMVEDFTVLCNIPY 134  
QY 134 MGEVVDQDGTFRILIKYNDGVHGDRE---CGRTINEIFVELNVALGVYDDDD--- 185  
135 MGEVVDQDGTFRILIKYNDGVHGDRE---CGRTINEIFVELNVALGVYDDDD--- 185  
QY 135 MGEVVDQDGTFRILIKYNDGVHGDRE---CGRTINEIFVELNVALGVYDDDD--- 185  
135 MGEVVDQDGTFRILIKYNDGVHGDRE---CGRTINEIFVELNVALGVYDDDD--- 185  
QY 186 -----DDDDDDDEEREKOKJEDHRDKESSRPFRKFPEDKIFEAISSMFPDKGTA 237  
186 -----DDDDDDDEEREKOKJEDHRDKESSRPFRKFPEDKIFEAISSMFPDKGTA 237  
QY 195 NDSIDGKODSKEDLPVTRKRNHAEKGNK--KSSK--KQFPDMIFSLASMFENGVP 250  
195 NDSIDGKODSKEDLPVTRKRNHAEKGNK--KSSK--KQFPDMIFSLASMFENGVP 250  
QY 238 EELKRYKELTREQALPGALPEECTPNIDGNVAVSVPBOSLSHFPULFCRCRYTCFLH 237  
238 EELKRYKELTREQALPGALPEECTPNIDGNVAVSVPBOSLSHFPULFCRCRYTCFLH 237  
QY 251 DDKKERYBELTMSDPAALPPQCTPNIDGNVAVSVPBOSLSHFPULFCRCRYTCFLH 310  
251 DDKKERYBELTMSDPAALPPQCTPNIDGNVAVSVPBOSLSHFPULFCRCRYTCFLH 310  
QY 298 PPHATPVTYRKNTETALDNKPGPCYCHLKGAKKPAALTAERIKTPEKRGRRGR 357  
298 PPHATPVTYRKNTETALDNKPGPCYCHLKGAKKPAALTAERIKTPEKRGRRGR 357  
QY 311 PPHATPVTYRKNTETALDNKPGPCYCHLKGAKKPAALTAERIKTPEKRGRRGR 364  
311 PPHATPVTYRKNTETALDNKPGPCYCHLKGAKKPAALTAERIKTPEKRGRRGR 364  
QY 358 LPNNSRPSTPTIN-VLESKDTSDREBAGTETGGENNDKEBEKKDSTSSSSANSRCQT 416  
358 LPNNSRPSTPTIN-VLESKDTSDREBAGTETGGENNDKEBEKKDSTSSSSANSRCQT 416  
QY 365 HATVSASCSNASSAVALTEBGSDDTG-----NDWASSSSSANSRCQT 409  
365 HATVSASCSNASSAVALTEBGSDDTG-----NDWASSSSSANSRCQT 409  
QY 417 PIRKKNPIBP-----ENVEWGAERAMFRVILIGTYNDFCAIARLIGTKCRQVY 467  
417 PIRKKNPIBP-----ENVEWGAERAMFRVILIGTYNDFCAIARLIGTKCRQVY 467  
QY 410 PTKOKASPAFPQCVVAPSEPVEMTGAESESLFRVFHGTGFNNFCSIARLIGTKCRQVY 469  
410 PTKOKASPAFPQCVVAPSEPVEMTGAESESLFRVFHGTGFNNFCSIARLIGTKCRQVY 469  
QY 468 EPRVYKSSITAPAPAEVNDTPPKTKRKRLMAHCRKIQLKDGSSNHYVYVOPCDHPR 527  
468 EPRVYKSSITAPAPAEVNDTPPKTKRKRLMAHCRKIQLKDGSSNHYVYVOPCDHPR 527  
QY 470 QFAVKE-SLILKQPTBELMTPSKRKRKRLMAHCRKIQLKDGSSNHYVYVOPCDHPR 528  
470 QFAVKE-SLILKQPTBELMTPSKRKRKRLMAHCRKIQLKDGSSNHYVYVOPCDHPR 528  
QY 528 QPCDSGCPVIAONFCEKFCQCSSECONRPFGRCKAQNTQCPCYLAARECDPLCLT 587  
528 QPCDSGCPVIAONFCEKFCQCSSECONRPFGRCKAQNTQCPCYLAARECDPLCLT 587  
QY 529 RPDSTPCLMTQWFCCKFCQCPD-----LARECDPLCLT 564  
529 RPDSTPCLMTQWFCCKFCQCPD-----LARECDPLCLT 564  
QY 588 CGAADHNDKNSVCKNSIGRSGKHLILAPSDVAGMIFIKDPVONNBFISYCGEIIIS 647  
588 CGAADHNDKNSVCKNSIGRSGKHLILAPSDVAGMIFIKDPVONNBFISYCGEIIIS 647  
QY 565 CGASEHNDKNSVCKNSIGRSGKHLILAPSDVAGMIFIKDPVONNBFISYCGEIIIS 624  
565 CGASEHNDKNSVCKNSIGRSGKHLILAPSDVAGMIFIKDPVONNBFISYCGEIIIS 624  
QY 648 QDEADRGRKYYDKMCSFLPNLNDPVLVADTRGKNIRFANHSVNPICVAKV 699  
648 QDEADRGRKYYDKMCSFLPNLNDPVLVADTRGKNIRFANHSVNPICVAKV 699





PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
XX Claim 1; SEQ ID NO 72833; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 880 AA;  
  
Query Match 18.9%; Score 719.5; DB 8; Length 880;  
Best Local Similarity 27.6%; Pred. No. 9.6e-46;  
Matches 231; Conservative 110; Mismatches 256; Indels 241; Gaps 38;  
  
25 RLRLQKRRARRADEVSKM---PSSNRKILERTILNQMKQRRIQPVHILTSVSLNGTR 81  
Db 59 RLSELKRIQGERVRSIKERANKKV-----DANVSP---PSSAASSRATA 103  
Qy 82 ECSVTSDDLPPPTQVIPLKTNANAVASVPIMYSWSPLOQNFVDETVLNT-----P 132  
Db 104 EDNGSNMNLSSRMWPLCKLNG-----FSHGVDSDRDYVPLKOVISASVXLPIAERIP 156  
Qy 133 YMGEBVLDDGTFFELIKATYDGKYNREGCFINDEFVLVNALGQYND--DDDDG 190  
Db 157 YTWIFLDRNQMAE-----DOSVGR--QIYYE-----QHGETLICSDE 197  
Qy 191 DPEEREKODLEDHDDKESRPPKPPSKIFFAISMFPDKGTAEELK----- 241  
Db 198 EEPERBEKREFSEB-----EDSIIMLIGER---GMEGEVODALCOLLSV 240  
Qy 242 -----EKYKELTEQQLFGALPPECTPNIDGPNAX-SVQRQ---SIHSFTLFCRCF 290  
Db 241 DASDLIERYNELKMD-----KQNTSEFSNGSFGDLISLKGJGALDSDNDFCRCL 294  
Qy 291 KYDCTLHPHATPNTYRKKN---TETALDNKPCGPGCQOHLGAKG-----PAA----- 336  
Db 295 VFDCLHGCQPLISASEKQPYWSDYEGDRKPCSHGCTQLQLAVAEVPETCSNFASKAE 354  
Qy 337 -----ALTAER---IKTPPKRPQ-----GRR----- 355  
Db 355 KASEBECKAVSVDVPHAAAGSVISQVETKDIGINDSSSGVEBDEHRIKREVPILKD 414  
Qy 356 -GRLENNSSRP-----STPTINLVESKDTDSREACTETGEGNNDKEEEK 400  
Db 415 SNDLPNLNKKQKTAADTKMSFVNSVPSLD--QALDSTKGQGGGTINKKVRDSEADAK 472  
Qy 401 K-----DTSSSSEANSRCQTP-----IKKKEPIPEPENTBMSGAESMRVLIG 445  
Db 473 EVGEBPIPNVNSYHDGSSICQPHHSGNGALIIAEMSETSRP--STEMNPTEKDLVLKGE 530

Qy 446 TYDNFCAIAR--LIGTKTCROYEF-RVKESSIAPD-----PAEVD-TP 488  
Db 531 IFGNSCLTARNLISGLKTCLDVSNYMRNEVSVFRRSSTPILLDDGRTDPNDDEVP 590  
Qy 489 PR-----KKRKRHLW-----AAHCKRIQAKDGSNNHYVNOPCDHPRQPCSSGCPVI 538  
Db 591 PRRLPRRKKTIRKLYSTKSASHPVSWKRIAGSKQSKQYPCG-CLSMCKDCPCLT 649  
Qy 539 AONFCEKFCQCSSECONRFPQRCR-KAQCNTKQCPCTLAWECDPDLCTGADHWDSK 597  
Db 650 NBTCEKTYCCCSKSCNRFRGCHCAKSCRSRQCPCEAAGREDDPVCRC-----W--- 701  
Qy 598 NVSCNCSI---QRG-----SKQHLILAPSDVAGNGIFIKDPVQNEFISEY 641  
Db 702 -VSGDGSILGEADPRRGGQGNRRLLRQQORILLGSDVAGAFGLKNSVSHREYLGEBY 760  
Qy 642 CGEIIISDDEADRGRKYDYDKMCSFLPYLANDPVTADRKNKIRPANHNVNPGYAKV 699  
Db 761 TGEIISHHEADKRGKTYDRANSSFLPDLDQVYLDQRKDKLFPANHSKAPNCYAKV 818  
  
RESULT 12  
ADT57579  
ID ADT57579 standard; protein; 895 AA.  
XX  
AC ADT57579;  
XX  
DT 13-JUN-2005 (first entry)  
XX  
DE Plant polypeptide, SEQ ID 7656.  
XX  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
OS Viridiplantae.  
XX  
XX US2004216190-A1.  
XX  
PD 28-OCT-2004.  
XX  
PF 18-DEC-2003; 2003US-00739930.  
XX  
PR 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.  
XX  
PA (KOVA/) KOVALIC D K.  
XX  
PI Kovalic DK;  
XX  
XX WPI; 2004-757369/74.  
XX  
XX Claim 2; SEQ ID NO 7656; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle



plant, to produce a transgenic plant. Promoters derived from the Me21 or Me22 genes are useful to direct expression of heterologous sequences. The modified variants of the PG genes are useful for generating or selecting antibodies immunoreactive to the non-variant polypeptide and to follow the segregation of chromosome regions and in marker assisted selection methods for crop improvement. Me21 and Me22 proteins can be used to repress the expression or prevent the expression of a desired target gene in specific tissue in a plant in vivo, to regulate homeotic gene expression in plants to create novel plants having improved agronomic traits. Protein derived from a PG gene of the invention is useful as an immunogen for producing monoclonal or polyclonal antibodies which are useful for screening plants for the expression of a Me21 or Me22 polypeptide, and for affinity chromatography for isolating a Me21 or Me22 polypeptide. This sequence represents the protein of Me22 of the invention

XX Sequence 893 AA;

SQ Query Match 18.3%; Score 698; DB 5; Length 893;

Best Local Similarity 25.7%; Pred. No. 4.3e-44; Mismatches 240; Gaps 42;

```
Matches 240; Conservative 103; Mismatches 240; Indels 352; Gaps 42;

OY 3 QTKKSEKRG-----PYCWRKRYKSEYMKRLQKFRABVKSMSBSNOKLE 51
DB 12 QRSKRSDDGMDGMDAAAASVPI-----HANTLTL-----IKVQSGRLAYIKE 54
OY 52 RTEILNOMKORR---IQVHILTSVSLRG-----RECVTSPL----- 89
DB 55 KLEV-NRKTLOHSSSLPVDVAAAYAS-RGTDGGMALSGRAAEQCG--SLANGIGER 110
OY 90 -----DPTQVPIPLKTLNAVA-----SVPIYMSVSPLOON-FWEDEET 126
DB 111 DVSVSHBENTAGTLATSSGATAQRTIVRFVKLPLVEKIPPTTWIFLIDKKQRMADDS 170
OY 127 VL--HNIPY--MGDEVL--DODGTPIRE-----LIRKYDQKV-----HEDRCGPTIN 167
DB 171 VGRBRRIYDVTGNEHALICSDSEDEIPEPEERKHFTKGEDHLIRAFQDHC-----LN 224
OY 168 DEIFPELVNAGVYDDDDDDDDGDDPEERERKOKLDEHDKESRPPKFPSP-----DKI 223
DB 225 QEV---VNVLCQF-----IGATPSREIERSVLF-RKMKHGSSSKIKSRSLDXT 272
OY 224 PEAISSMPEDKGTAELEKRYELTEQQLPGALPECTPNIIDGPAKSVQREQSLSHPTT 283
DB 273 MDV-----LDSEFVN 282
OY 284 LFCRCFKTKDCFLH-----PHATPNTYKKNQETALDNKRCGPQY-----Q 326
DB 283 LFCRCFLVDCRIHGCSQNLVPPCEKQPYSPDENK-----KCGHLCYLRFQ 332
OY 327 HLEGAKE----- 333
DB 333 WREGFKEHDDGLAGATTMESGTRASRVNVNMTRESDSNRKQKNTIRSMVLVGSBK 392
OY 334 FAALTAERIKTPP-----KPGGRRRGRLPNNSSRSPSTPTI 370
DB 393 IISVSASBESTTPADISETENVSDDLPPSLRKHKISGHGGRYRHSBGKQKQVFTSDI 452
OY 371 -----NVLESQDTSDRBACTRGCGENNDKEBKQDETSSSSAANRC-QTPIK 419
DB 453 SFEQSIMNQLSIPKIRDT---RLBSRESG---DKRLILBISTYKTSRKMKCGSSPPT 504
OY 420 MKPNLEPPEN-----VE-----MSGASAFRFLIGTYDNFCALAR--LIGTKTCR 464
DB 505 TMEVNGROSNKYSTGNFLESTLSCSALERDLYKIGITIPKNGSLIRANLISGLKTCI 564
OY 465 QYEFERVKESSIIAPAP-----AEDVDTPEPKKKRRR-----LW- 499
DB 565 EVANVTYNNAGAAAKKPLINKSISGDPFAENEDQYMEQDMAARTIRYRRGRNRKLYTKW 624
OY 500 -AAHCKRIQLKDGSSNNHYNTQPCHPQPCDSSPCVYIAQNFCEKFPQCGSSQCONRFP 558
DB 625 SAGH-PTVAKRTDDGQCYTQYSPC-ACQOMCGKDCPCADKGTCCCKYGGCSKSKNKR 682
```

```
OY 559 GCRC-KAQCNTKQCPCYLAVERCDPDLCTRGADHDS-----KNVSCNKC 604
DB 683 GCHCAKQCRSRQCPCPAASRECDPVCRCN-----WVSCDGSIGEPRLARGDGYQCGNM 737
OY 605 SIQSGSKKHLIAASDVAGWCIFFIKDPVQKNEFISEYCGEIIQDDEADRGKYYDKWCS 664
DB 738 KLLKKQQRILLGSDVAGWGAFFIKNPVNRNDYGEYTGELISHKQADKRGKLYDRANS 797
OY 665 PLFNLNDPVVDATRKGNKIRFANHSVNPNCYATV 699
DB 798 PLFNLNDGYVDATRKGDKLFANHSNPNCYATV 832
```

RESULT 14

ADM48181

ID ADM48181 standard; protein; 891 AA.

AC ADM48181;

DT 03-JUN-2004 (first entry)

XX Polypeptide sequence #231 useful in producing transgenic plants.

XX Plant; maize; rice; sorghum; thale cress; soybean; wheat; transgenic;

XX osmotic stress; sugar transport; cell cycle pathway; plant height;

XX carboxylate transport; crop productivity; plant growth;

XX stress resistance; disease resistance; insect resistance; heat tolerance;

XX nitrogen assimilation; water stress tolerance;

XX photosynthetic carbon fixation; virus resistance; gene therapy.

OS Zea mays.

PN US2003233670-A1.

PD 18-DEC-2003.

PF 04-DEC-2002; 2002US-00310154.

PR 04-DEC-2001; 2001US-0337358P.

PA (EDGE/) EDGERTON M. D.

PA (CHOM/) CHOMET P. S.

PA (LACC/) LACCETTI L. B.

PI Edgerton MD, Chomet PS, Laccetti LB;

PT MPI: 2004-061374/06.

DR N-PSDB; ADM47813.

XX Claim 8; SEQ ID NO 599; 144bp; English.

PT The present invention relates to polynucleotide sequences, and the

PT proteins they encode. The sequences are isolated from a variety of

PT organisms such as plants (e.g. maize, rice, sorghum, thale cress,

PT soybean, and wheat), cyanobacteria, bacteria, yeast and other fungi. The

PT polynucleotide and polypeptide sequences that have improved properties. Also

PT disclosed are methods of producing fertile transgenic plants, preferably

PT maize, with desired phenotypes. The polynucleotide and polypeptide

PT sequences are useful for improving plants by providing protection against

PT osmotic stress, improving altering sugar transport and/or metabolism,

PT modifying the cell cycle pathway, reducing plant height, modifying

PT carboxylate transport, improving crop productivity, improving plant

PT growth and stress resistance, improving disease resistance, improving

PT insect resistance, improving cold or heat tolerance, improving nitrogen

PT assimilation, improving stalk strength, improving water stress tolerance,

PT improving photosynthetic carbon fixation, improving biotic and abiotic

PT stress resistance, improving resistance to oxidative stress, providing

CC Increased vigour, reducing senescence, and conferring virus resistance.  
CC The present sequence represents a polypeptide sequence of the invention.  
CC Note: The sequence data for this patent is not provided in the printed  
CC specification but is obtained in electronic format from the USPTO website  
CC at [seqdata.uspto.gov](http://seqdata.uspto.gov).  
CC  
XX

SQ Sequence 891 AA;

Query Match 18.2%; Score 693; DB 8; Length 891;  
Best Local Similarity 25.6%; Pred. No. 1e-43;  
Matches 239; Conservative 103; Mismatches 241; Indels 350; Gaps 42;

```
QY 3 QGKSKSEKQ-----PVCGRKRVSEYMRQLKFRADRYKMSFSSNRKQILE 51
DB 12 QSKSKDDQMGKDAALASVPI-----HANLQL-----IKVQSGRLAYIRE 54
QY 52 RTEILNQEWKOR---IQPVHILTVSSSLGT-----RECSVTSDL----- 89
DB 55 KLEEV-NRKTILQRISCSLFDVAALAAVVAS-RGTGGNALSGRAARQCG--SDLANGIGER 110
QY 90 -----DPTVOITFLKTLNANA-----SVPIYMSVPLQON-FWVEDET 126
DB 111 DVVSVQEBMLATGTLALSSGAPARTIVFVFLPLVEKIPYTTWIFLDKQRMADDS 170
QY 127 VL--HNIPY--MGDEVL---DODGTLEB-----LIKNYDGKV-----HGDREGCFIN 167
DB 171 VGGRRRIYDYVGNELICSDSDDEIPEPEBEKHPTKGEDHLIWRATODHG-----LN 224
QY 168 DRIFVELVNALGQYNDDDDDDDDEBEREKOLEDHRDDKESRPRKFPs---DKI 223
DB 225 QEV---VAVLCQF-----IGATPSIEIRSEVLF-KNEHGSSSDKIEIRSLDXT 272
QY 224 FRAISMPDKGTAELEKEKYELTEQLPGALPECTPNIDGPAKAVQREQSLHSPT 283
DB 273 MDVAV-----LDSFDN 282
QY 284 LPCRRCFKYDCPLH-----PFAHPNTYKRRNTETALNKPQGPQY-QHLE 329
DB 283 LPCRRLVDFCDRLHGCSQNLVPPCEKQFYSFDENK-----KCGHLCYLNRMB 332
QY 330 GAKK-----FAA 336
DB 333 GPEKMHDDGLAGATYTMESGTASQRVDVNVVSHESDSNRKKNISMTLVGSSKIIS 392
QY 337 ALTAERIKTPP-----KPPGRRRRGLPNNSSRPSTPTI-- 370
DB 393 SVASBESTTPPSADISSETENVSSDLPPSLRKHKISKHGPRYREHSPGRQKVFISDISF 452
QY 371 -----NVLSKQDSDREAGTENGNNDEBEKKDETSSSSEANSRC-QTPIYMK 421
DB 453 EENIMNKLISIPETIDT--RLBSRESGG-----DKRLIIDESEYKTSRKDMCGESPATTM 504
QY 422 PNIEPPEN-----VSGABASMPFVLIGTYVDFCAIAR--LIGTKICROV 466
DB 505 ENVGQSNVYSTKYNLESTLSCMSALREDDYIKGIRIGKISCLIAKMLISGLKTCIEV 564
QY 467 YEFVRESISIAAP-----AEDVDTPPRKKRKR-----LM-A 500
DB 565 ANVMYNNPAMAKRPLINKSISGDFALNEQDYMEQDMARTRIYRRGRNRKRLKYTWKSA 624
QY 501 AHCRKQLKKGSSNNVNYVQPCDHRQPCDSSCPVIAQNPCKECCOSSSCQNRFPFC 560
DB 625 GH-PTVRRKTDGKQCTQYSFC-ACQQWCGKDCPADKGTCCBKCGSKCKXNFRSC 682
QY 561 RC-KAQCNTKQCPCYLAVERCDPDLCTLTCGADHMS-----KNVSCNKNCI 606
DB 683 HCAKSGCRSRQCPFAASRECDPDVCRNC-----WVSCDGSIGREPLANGDVGCCNMTL 737
QY 607 QRGSKKHLLADSDVAGMGIPIKDPVQKNPFISEYCGEIIISQDEADRGRKYDKWCSFL 666
DB 738 LKQOQRILLGSRSDVAGMGAFLIKNPVKNQDYLGEYGLISHKADKRKGIYDRANSPL 797
QY 667 FPLNNDPVDATRKNGKIRFANHSVNPNCYAY 699
```

DB 798 FPLNDQYVDATRKNGKIRFANHSNPNCYAY 830

RESULT 15  
ADT57263  
ID ADT57263 standard; protein; 931 AA.

AC ADT57263;

DT 13-JAN-2005 (first entry)

XX Plant polypeptide, SEQ ID 7340.

XX plant; transgenic; cold tolerance; growth rate; drought tolerance;  
XX disease resistance; galactomannan production; plant growth regulator;  
XX heat tolerance; herbicide tolerance; lignin production;  
XX extreme osmotic condition tolerance; pathogens resistance;  
XX pest resistance; yield improvement; seed oil yield; seed protein yield.

XX Viridiplantae.

XX US2004216190-A1.

XX 26-OCT-2004.

XX 18-DEC-2003; 2003US-00739930.

XX 28-APR-2003; 2003US-00424599.

XX 28-APR-2003; 2003US-00425115.

XX (KOVA/) KOVALIC D K.

XX Kovalic DK;

XX WPI; 2004-757369/74.

PT New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.

XX Claim 2; SEQ ID NO 7340; 14pp; English.

XX The invention relates a recombinant DNA construct comprising a  
XX polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO:  
XX 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
XX (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,  
XX Arabidopsis, wheat and rape but the specification does not indicate which  
XX sequences is derived from which organism. Also included is a method of  
XX producing a plant having an improved property, comprising transforming a  
XX plant with a recombinant DNA construct comprising a promoter region  
XX functional in a plant cell operably joined to a polynucleotide encoding a  
XX polypeptide associated with the property, and growing the transformed  
XX plant. The property is selected from improving plant cold tolerance, for  
XX manipulating growth rate in plant cells by modification of the cell cycle  
XX pathway, for improving plant drought tolerance, for providing increased  
XX resistance to plant disease, for galactomannan production, for production  
XX of plant growth regulators, for improving plant heat tolerance, for  
XX improving plant tolerance to herbicides, for increasing the rate of  
XX homologous recombination in plants, for lignin production, for improving  
XX plant tolerance to extreme osmotic conditions, for improving plant  
XX tolerance to pathogens or pests, for yield improvement by modification of  
XX photosynthesis, for modifying seed oil yield and/or content, for  
XX modifying seed protein yield and/or content, for yield improvement by  
XX modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
XX and for yield improvement by providing improved plant growth and  
XX development under at least one stress condition. The polynucleotide may  
XX also encode a plant transcription factor. The methods and compositions of  
XX the present invention are useful in the field of biochemistry and  
XX genetics, in particular for producing transgenic plants with improved  
XX biological characteristics such as increased yield, improved nitrogen  
XX flow, increasing plant tolerance to cold or heat, improving plant  
XX tolerance to extreme osmotic and drought conditions, and improving plant



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:21:54 ; Search time 27.573 Seconds  
(without alignments)  
2439.179 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808

Sequence: 1 MGQTKSKSEKGPVCMKRRVK.....KGNKIRFANHSVNPVCYAKV 699

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 80:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3802	99.8	746	2	enhancer-of-zeste
2	719.5	18.9	856	2	polycomb protein B
3	704.5	18.5	898	2	hypothetical protein
4	679.5	17.8	902	2	curly leaf protein
5	612	16.1	689	2	protein MEDA [imp
6	519	13.6	775	2	hypothetical prote
7	228.5	6.0	920	2	probable transcrip
8	228.5	6.0	1080	2	hypothetical prote
9	228	6.0	3968	2	trithorax homolog
10	226.5	5.9	2144	2	ash1 protein - fru
11	226	5.9	3869	2	protein C26B.12 l
12	220.5	5.8	1802	2	unknown protein F2
13	219	5.8	528	2	hypothetical prote
14	210	5.5	408	2	protein C26B.10 l
15	208.5	5.5	739	2	hypothetical prote
16	205	5.4	1767	2	hypothetical prote
17	204	5.4	2588	2	NSD1 protein - mou
18	201.5	5.3	788	2	similar to mammai
19	200	5.3	635	2	Su(var)3-9 protein
20	198	5.2	1001	2	G9a protein - huma
21	195.5	5.1	352	2	hypothetical prote
22	193.5	5.1	203	2	probable SET-domai
23	192	5.0	344	2	hypothetical prote
24	190.5	5.0	490	2	clt4 protein - fis
25	190.5	5.0	490	2	mating-type loci a
26	188.5	5.0	3828	2	trithorax protein
27	188	4.9	3759	2	trithorax protein
28	185.5	4.9	794	2	similar to mammai
29	184	4.8	477	2	hypothetical prote

30	179	4.7	733	2	hypothetical prote
31	176	4.6	898	2	hypothetical prote
32	175.5	4.6	798	2	hypothetical prote
33	175	4.6	954	2	hypothetical prote
34	174.5	4.6	4957	2	ALR protein - huma
35	174.5	4.6	5262	2	hypothetical prote
36	172	4.5	669	2	hypothetical prote
37	166.5	4.4	1193	2	ALR protein homolo
38	162.5	4.3	2422	2	neurofilament prot
39	161	4.2	1200	2	protein C43B1.3 l
40	161	4.2	1590	2	erythrocyte membra
41	160.5	4.2	1729	2	hypothetical prote
42	160	4.2	401	2	hypothetical prote
43	159	4.2	912	2	probable SET-domai
44	158	4.1	186	2	functional sarcopl
45	155	4.1	706	2	

#### ALIGNMENTS

##### RESULT 1

G02838

enhancer-of-zeste homolog 2 - human

C.Species: Homo sapiens (man)

C.Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 09-Jul-2004

C.Accession: G02838

R.Antonarakis, S.B.

submitted to the EMBL Data Library, February 1996

A.Reference number: H01746

A.Accession: G02838

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-746 <ANT>

A.Cross-references: UNIPROT:Q15910; UNIPARC:UPI000016B348; EMBL:X95653; NID:G1438063

A.Genetics:

A.Gene: GDB:EH2; EZH1

A.Cross-references: GDB:701613; OMIM:601573

A.Map position: 21q22.2-21q22.2

Query Match	99.8%	Score 3802;	DB 2;	Length 746;
Best Local Similarity	99.8%	Pred. No. 2.6e-222;		
Matches 698;	Conservative	0;	Mismatches 1;	Indels 0;
QY	1	MGQTKSKSEKGPVCMKRRVKSRRLRLQKFRPADYKSMFSSNRKILERTILNQEW	60	
DB	1	MGQTKSKSEKGPVCMKRRVKSRMYRLRLQKFRPADYKSMFSSNRKILERTILNQEW	60	
QY	61	KORRIQPVHILTSVSLGTRSCVTSDDLDPPTQVYIPLKTLNAAVASVPIMTSMSPLOQNF	120	
DB	61	KORRIQPVHILTSVSLGTRSCVTSDDLDPPTQVYIPLKTLNAAVASVPIMTSMSPLOQNF	120	
QY	121	MWEDTVLHNIPLYMGDEVLDQGTPIEBLIRKYDVKVAGDBRCGFINDEIFELVNALGQ	180	
DB	121	MWEDTVLHNIPLYMGDEVLDQGTPIEBLIRKYDVKVAGDBRCGFINDEIFELVNALGQ	180	
QY	181	YNDDDDDDDGDDPERRERKQKOLEBHRDQKSRPPKPSDKIFPAISMPFDKGTABRL	240	
DB	181	YNDDDDDDDGDDPERRERKQKOLEBHRDQKSRPPKPSDKIFPAISMPFDKGTABRL	240	
QY	241	KEKYKELTEOOLPGALPEECTPNIDGPNAKSVORQSLHSFHTLFCRRCFKDKCLAHPPH	300	
DB	241	KEKYKELTEOOLPGALPEECTPNIDGPNAKSVORQSLHSFHTLFCRRCFKDKCLAHPPH	300	
QY	301	ATPNTYKRNKTEALDNPCGPOCYOHLGAKEPAAALTAIRIKTPPKRPGRRGRRLPN	360	
DB	301	ATPNTYKRNKTEALDNPCGPOCYOHLGAKEPAAALTAIRIKTPPKRPGRRGRRLPN	360	
QY	361	NSRPSPTTIVLBSKQSDSREACTGTGGRNNDKEERKQDEFSSSEANSRCQTPYK	420	
DB	361	NSRPSPTTIVLBSKQSDSREACTGTGGRNNDKEERKQDEFSSSEANSRCQTPYK	420	
QY	421	KPNIEPPNVMMSGABMFRVLIGTYVDNFCALRLGTGTCRCQVVEFRVASESIIADA	480	
DB	421	KPNIEPPNVMMSGABMFRVLIGTYVDNFCALRLGTGTCRCQVVEFRVASESIIADA	480	

```
Db      421  |||||
      421  KPNIEPPNENWMSGABMFVLTIGTYDNCFAIRLIGTKICRVYFFRYKSESSIIAPA 480
      481  PABVDTPPRKKGRHRLMAAHCRIQLKDGSSNHVNYNQCCHPROPCSSCPCTVAAQ 540
      481  PABVDTPPRKKGRHRLMAAHCRIQLKDGSSNHVNYNQCCHPROPCSSCPCTVAAQ 540
      481  PABVDTPPRKKGRHRLMAAHCRIQLKDGSSNHVNYNQCCHPROPCSSCPCTVAAQ 540
      541  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      541  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      541  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      541  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      601  CKNCISQSGSKHLLASDVAGWGIPIKDPVQKNEFISEGCIISQDEADREKGYDX 660
      601  CKNCISQSGSKHLLASDVAGWGIPIKDPVQKNEFISEGCIISQDEADREKGYDX 660
      601  CKNCISQSGSKHLLASDVAGWGIPIKDPVQKNEFISEGCIISQDEADREKGYDX 660
      601  CKNCISQSGSKHLLASDVAGWGIPIKDPVQKNEFISEGCIISQDEADREKGYDX 660
      661  YMCSEFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
      661  YMCSEFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
      661  YMCSEFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
      661  YMCSEFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
```

## RESULT 2

T52415

polycomb protein EZA1 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear crease)

C:Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 09-Jul-2004

C:Accession: T52415

R:Biodeau, P.; Luo, M.; Dennis, E.S.; Peacock, W.J.; Chaudhury, A.M.

Submitted to the EMBL Data Library, October 1998

A:Reference number: Z26069

A:Accession: T52415

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-856 &lt;BIU&gt;

A:Cross-references: UNIPROT:Q9ZSM8; UNIPARC:UPI0000A5399; EMBL:AF100163; PIDN:AND09108.

Query Match 18.9%; Score 719.5; DB 2; Length 856;

Best Local Similarity 27.6%; Pred. No. 66-36;

Matches 231; Conservative 110; Mismatches 256; Indels 241; Gaps 38;

```
      25  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      25  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      35  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      35  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      82  ECSVTSDDLDPFTQVYIPLKTLNANVASVPIMYSWSPLOQNFWEDETVLHNI-----P 132
      82  ECSVTSDDLDPFTQVYIPLKTLNANVASVPIMYSWSPLOQNFWEDETVLHNI-----P 132
      80  EDNGNSNMLSSRMWPLCKLNG-----FSHGVDGRDYVPTKOVISASVKLPIAEKIRP 132
      80  EDNGNSNMLSSRMWPLCKLNG-----FSHGVDGRDYVPTKOVISASVKLPIAEKIRP 132
      133  YMGDEVLDQDGTFFIELIKNYDGKYNHGRBEGCFINDEIFVELVNALGQYNDND--DDDG 190
      133  YMGDEVLDQDGTFFIELIKNYDGKYNHGRBEGCFINDEIFVELVNALGQYNDND--DDDG 190
      133  YTMWIFLDRNQMAE-----DQSVVGR-----QIYYE-----QHGEETLIGDSE 173
      133  YTMWIFLDRNQMAE-----DQSVVGR-----QIYYE-----QHGEETLIGDSE 173
      191  DDEPEREKQDLERHDDKESRPPKRPSPDKIFPAISMPDPDKTAEELK-----241
      191  DDEPEREKQDLERHDDKESRPPKRPSPDKIFPAISMPDPDKTAEELK-----241
      174  EEPEREEKERFESG-----EDSIITWIGQY-----GMGEFVQDALCOLLSV 216
      174  EEPEREEKERFESG-----EDSIITWIGQY-----GMGEFVQDALCOLLSV 216
      242  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      242  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      217  DASDILERYNELKLDK-----KQNTBEPNSNGFGLGISLEKIGALALDSFDNLFCRCL 270
      217  DASDILERYNELKLDK-----KQNTBEPNSNGFGLGISLEKIGALALDSFDNLFCRCL 270
      291  KYDCFLH--PHTATNTYKKNK--TETALDNKPCGQOCTQHLBGAKE-----PAA-----336
      291  KYDCFLH--PHTATNTYKKNK--TETALDNKPCGQOCTQHLBGAKE-----PAA-----336
      271  VFDCRLHGCQPLISASERKQPYMSYBDRKPCSGHCTIQLKAVAEVETCSNFAKAE 330
      271  VFDCRLHGCQPLISASERKQPYMSYBDRKPCSGHCTIQLKAVAEVETCSNFAKAE 330
      337  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      337  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      331  KASERCKAVASDVPHAAASGSLQVEKTDIGIANDVSSSGVEQEGHIGRKEKRPILKD 390
      331  KASERCKAVASDVPHAAASGSLQVEKTDIGIANDVSSSGVEQEGHIGRKEKRPILKD 390
      356  -GRLENNSSRP-----STPTINYLESKDTPSDREACTETGEGENNDKEBEK 400
      356  -GRLENNSSRP-----STPTINYLESKDTPSDREACTETGEGENNDKEBEK 400
      391  SNDLPNLNKKQKTAADTKMSFVNSVPSLD--QALDSTKDGQGTTONKNVRBDEADAK 448
      391  SNDLPNLNKKQKTAADTKMSFVNSVPSLD--QALDSTKDGQGTTONKNVRBDEADAK 448
      401  K-----DTSSSSEANSKQCF-----IKKKPNIEPPNENWMSGABMFVLTIG 445
      401  K-----DTSSSSEANSKQCF-----IKKKPNIEPPNENWMSGABMFVLTIG 445
```

```
Db      449  EVGEPIFDSNVHDSGSSICQPHHSGNGAIIIMESTSRP--STENWPIEKOLYLKGE 506
      446  TTYDNPFAIAR--LIGTKTCROYER--RYVESIIAPA-----PAEDVD--TP 488
      446  TTYDNPFAIAR--LIGTKTCROYER--RYVESIIAPA-----PAEDVD--TP 488
      507  IFGRNSCLIANRNLISGLTKLGDVSNVYKREHVSIFRNSSTPNNLLDDGRTPGDNDNEVP 566
      507  IFGRNSCLIANRNLISGLTKLGDVSNVYKREHVSIFRNSSTPNNLLDDGRTPGDNDNEVP 566
      489  PR---KKKAKHLM-----AAHCRIQLKDGSSNHVNYNQCCHPROPCSSCPCTVAAQ 538
      489  PR---KKKAKHLM-----AAHCRIQLKDGSSNHVNYNQCCHPROPCSSCPCTVAAQ 538
      567  PRTLPFRKGRKTRDLKSTYSAGHPVSWKSIAGKNQSCKQYPCG--CLSMGDCDPCLT 625
      567  PRTLPFRKGRKTRDLKSTYSAGHPVSWKSIAGKNQSCKQYPCG--CLSMGDCDPCLT 625
      539  AONFCERKPCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 597
      539  AONFCERKPCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 597
      626  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      626  NPEKFCQCCSSBQONRPFQCRCKAQACNTKQCPCTLAVERCDPDLCTLTCGAADHMSKNVS 600
      598  NVSCKNCISI---ORG-----SKHLLASDVAGWGIPIKDPVQKNEFISEY 641
      598  NVSCKNCISI---ORG-----SKHLLASDVAGWGIPIKDPVQKNEFISEY 641
      678  -VSCGDGSLGEPAPRGRGQCGGNMRLLRQQRILIGKSDVAGWGAFLKNSVSKNEYLGEY 736
      678  -VSCGDGSLGEPAPRGRGQCGGNMRLLRQQRILIGKSDVAGWGAFLKNSVSKNEYLGEY 736
      642  CGEITISQDEADREKGYDXKTCMGSFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
      642  CGEITISQDEADREKGYDXKTCMGSFLPNLNDPVVDATRKNGKIRFANHNVNPNCTACV 699
      737  TGBLISHHREADKRGKIDYDRANSFPLFDLNDQYVLDQKDKLKFANHSKAPNCTACV 794
      737  TGBLISHHREADKRGKIDYDRANSFPLFDLNDQYVLDQKDKLKFANHSKAPNCTACV 794
```

## RESULT 3

T101503

hypochemical protein T10M13.3 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear crease)

C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999

C:Accession: T01503

R:Johnson, A.P.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; Got

Martensen, R.; McComble, W.

Submitted to the EMBL Data Library, May 1997

A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.

A:Reference number: Z14346

A:Accession: T01503

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-898 &lt;JOH&gt;

A:Cross-references: UNIPARC:UPI0000A0AC3; EMBL:AF001308; NID:g2104523; PID:g3912918

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4S

A:Inserts: 48/2; 83/1; 118/3; 207/2; 260/2; 288/3; 336/3; 352/3; 554/2; 603/3; 677/2; 7

Query Match 18.5%; Score 704.5; DB 2; Length 898;

Best Local Similarity 27.9%; Pred. No. 51e-35;

Matches 228; Conservative 108; Mismatches 259; Indels 221; Gaps 38;

```
      25  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      25  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      101  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      101  RLRLQKFRFRADVEYKSM--PSSNRQKILERTETILNOEMKORRIQPVHILTSVSLRGTR 81
      82  ECSVTSDDLDPFTQVYIPLKTLNANVASVPIMYSWSPLOQNFWEDETVLHNI-----P 132
      82  ECSVTSDDLDPFTQVYIPLKTLNANVASVPIMYSWSPLOQNFWEDETVLHNI-----P 132
      146  EDNGNSNMLSSRMWPLCKLNG-----FSHGVDGRDYVPTKOVISASVKLPIAEKIRP 198
      146  EDNGNSNMLSSRMWPLCKLNG-----FSHGVDGRDYVPTKOVISASVKLPIAEKIRP 198
      133  YMGDEVLDQDGTFFIELIKNYDGKYNHGRBEGCFINDEIFVELVNALGQYNDND--DDDG 190
      133  YMGDEVLDQDGTFFIELIKNYDGKYNHGRBEGCFINDEIFVELVNALGQYNDND--DDDG 190
      199  YTMWIFLDRNQMAE-----DQSVVGR-----QIYYE-----QHGEETLIGDSE 239
      199  YTMWIFLDRNQMAE-----DQSVVGR-----QIYYE-----QHGEETLIGDSE 239
      191  DDEPEREKQDLERHDDKESRPPKRPSPDKIFPAISMPDPDKTAEELK-----241
      191  DDEPEREKQDLERHDDKESRPPKRPSPDKIFPAISMPDPDKTAEELK-----241
      240  EEPEREEKERFESG-----EDSIITWIGQY-----GMGEFVQDALCOLLSV 282
      240  EEPEREEKERFESG-----EDSIITWIGQY-----GMGEFVQDALCOLLSV 282
      242  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      242  -----EKYKELTEQQLRGALPPECTPNIDGNNAK--SVQRBQ-----SLHSFHTLFCGRCF 290
      283  DASDILERYNELKLDK-----KQNTBEPNSNGFGLGISLEKIGALALDSFDNLFCRCL 336
      283  DASDILERYNELKLDK-----KQNTBEPNSNGFGLGISLEKIGALALDSFDNLFCRCL 336
      291  KYDCFLH-----PHTATNTYKKNK--TETALDNKPCGQOCTQHLBGAKEFAAA--337
      291  KYDCFLH-----PHTATNTYKKNK--TETALDNKPCGQOCTQHLBGAKEFAAA--337
      337  VFDCRLHGCQPLISALKAVAEVETC--SNFASKAEKASERCKAVASDVPHAAASG 394
      337  VFDCRLHGCQPLISALKAVAEVETC--SNFASKAEKASERCKAVASDVPHAAASG 394
```



QY 338 ---LTAAR-----IKTPPKRG-----GRR-----GRLPNNSRP----- 365  
DB 395 VSLQREKTDIGIKNDSSSGVEQEHGIRKREVPILKSDNDPUNSNKQKTAASDTKMS 454  
QY 366 ---STPTINVLBSKOTDSREAGETGTGENDKEBEKK-----DETSSSSSEANSRCQTP 417  
DB 455 FVNSVPSLD--QALBSTKGDDGTTDNKYNRDSEADAKVGPPIPNDSVHDGSSSICPH 512  
QY 418 -----IOMKRIEPEPEVNSGAASMFVILGTYYNFCALAN--LIGTYTCQ 465  
DB 513 HSGNGGAILIIMSETSRP--STEMNPIEKDLYLKVGEI.FGRNSCLIANLLSGIKTCLD 570  
QY 466 VYEF-RVXESSIIAPA-----PAEDVD--TPPR-----KKGRGRLM-----A 500  
DB 571 VSNYREHEVSVFRSSSTNNLLDQKTDPGNDNDVPPRTLPFRKGRKTKLKTSTYSA 630  
QY 501 AHCRKIQLKQDSSNMHVNYOPCDHPROPDSSCPVLAQNFCERKQCCSSCONRFPQC 560  
DB 631 GHPVWKRIAGKNGSCQKYTEPCG--CLSNWCGHDCPLTNBCTCEKQCCGSKCNRRFC 689  
QY 561 RC-KAQCNTKQCPCTLAVERCDPLCTCGAADHWDKSNVSCNCSI---ORG----- 609  
DB 690 HCANKQCSRSQPCPAAGRECDPVCRCN-----W-----VSCDGSIGRAPRRGQCCGN 740  
QY 610 -----SKKHLIAPSDVAGWGI.FIKDPYQKNEPISEYCGELISQDEADRRGKYVDKMC 663  
DB 741 MELLIRQOORILKGSVDVAGWGAFLKNSVKNBYLGEYTGELISHHEDAKRKIKYDRANS 800  
QY 664 SFLFNLANDPVVDAFRKGNKIRFANHSVNPNCYAV 699  
DB 801 SFLFNLNDQYVLDAQRKDKLKFANHSANPCYAV 836

## RESULT 4

101127  
curly leaf protein (polycomb-group) [imported] - Arabidopsis thaliana  
N.Alternate names: curly leaf protein homolog F26B6.3  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Feb-2001  
C.Accession: T01127; A84624  
R.Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
submitted to the EMBL Data Library, June 1998  
A.Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.  
A.Reference number: Z14198  
A.Accession: T01127  
A.Status: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-902 <ROU>  
A.Cross-references: UNIPARC:UPI00000A76B5; EMBL:AC003040; NID:g3242700; PID:g3242729  
A.Experimental source: cultivar Columbia  
R.Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
euser, D.; Nierman, W.C.; White, O.; Eileen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J  
Nature 402, 761-766, 1999  
A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A.Reference number: A84420; MUID:20083487; PMID:10617197  
A.Accession: A84624  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-902 <STO>  
A.Cross-references: UNIPARC:UPI00000A76B5; GB:AB002093; NID:g3242729; PIDN:AA03781.1; C  
C.Genetic8:  
A.Gene: At2g23380; F26B6.3  
A.Map position: 2  
A.Introns: 21/3; 52/3; 153/2; 207/2; 235/3; 280/3; 296/3; 320/3; 554/2; 604/3; 679/2; 72  
Query Match 17.8%; Score 679.5; DB 2; Length 902;  
Best Local Similarity 25.5%; Pred.No. 1.7e-33;  
Match 230; Conservative 107; Mismatches 269; Indels 295; Gaps 35;  
QY 9 EKGPVCMKRVKSEYRRLQLKFRPADVKSMSFSSNKKILERILINQEWKRRIPV 68  
DB 24 ERGPA--SKVSEVIESLKKGLADRCISIKRIDENKNLPAITQSPMR----- 71

QY 69 HILTSVSLRGTRCSVTSDL-----DEPTQVILPLKTLN-----AVASV 107  
DB 72 -----SSMERGSSCKDSDLLVKRQRDS.PGKMSGIDSSNNRYYVEDGPASSGMVQGVSSV 125  
QY 108 PIMYSPLQOQFMWEDETVLNIPYMGDEVLDQGTFLBELIKNYDKGVHDBRCGFIN 167  
DB 126 PKYLSRLPIKM-----PDIKLSPTTYTWFLDRQORTB-----DSVGR-- 167  
QY 168 DEIFVELVNALQYNDDDDDGDDPBERBERKQDLIEDHDKESRPPRK-- 218  
DB 168 -RIYDQGGGALICSDSBEALDDEERK-----RDPLF--PEDYIRMTLRL 213  
QY 219 -PSDIFPAISGMFPDQKTABELKAYELTEQQLPGLPPECPTNIDGNPAKV---QR 274  
DB 214 GLSDSVLALAS-FLSRST--EIKARHVLNKK-----EVSSGNOQAESLINDM 264  
QY 275 EOSLSPFLLFCRRCFYKDCFLH-----PFAATPYTKRKTETALDNKPCGP 322  
DB 265 BGALDSFNLFCRCLVDFDCRLHGCSDLIIPAEKPAWCP-----PYDENULTGA 315  
QY 323 QCYOHLGAKAFPAALTAE-----RIKTPPKRGRRRLPNN--SSRSTPTI 370  
DB 316 NCYKTLKSGRPPGVGTIEGKTGSSDGAQKTTPTTKSSKLNKGRKPTFPESASSNK 375  
QY 371 NVLES-----KDTSD-----REAGTGTGENN----- 393  
DB 376 CALFTSDENGLQDPTNSDKVSSSPKVGSGRRVRKKNRVARVPRKTKOKKTEA 435  
QY 394 -----DKBERKKDERTSS-----SEANRCORPIPMKPY----- 423  
DB 436 SDSDSIASGSCSPSDAKKDNEDATSSQKAVKSGNSGKSKNGTFAVSNNSVQDVPV 495  
QY 424 -----IEPPENVE-----WSGABSMFVILGTYYNFC 452  
DB 496 COSNEVASBELAPGSDSELARKBFFMGTVSHRGLATNKLMPLEKSLPDKVGEIRGNMSC 555  
QY 453 AIAAR-LIGTYTCQVYFRV-----KESIIAPAPADVDTPPKKKRK 495  
DB 556 LIARNLISGFSCEVFPYMTCSBNKASFPFGDGLNPDGSSKRFIDNGMNVNQQVRRSRF 615  
QY 496 HR-----LW-----AAHC--RKLQKQDSSNMHVNYOPCDHPRO-----PCDS 532  
DB 616 LRRGKVRRLKYTWKSAHYHSIRKITEKKO-----OPC--RFPNPNCKIACGK 663  
QY 533 SCPCVIAQNFCKRCQCCSSSECONRFPFGRC--KAQCTKQCPCTLAVERCDPDLCTTCGA 591  
DB 664 ECPCLNLTGTCCKKTCGCKSKCNRRFRGCHCAKSCRSRQCPAPADRCDDPVCNCC-- 720  
QY 592 DHW-----DSKNVSCQKCSIQRGSKHLIAPSDVAGWGI.FIKDPYQKNEPI 638  
DB 721 --WYIGDGSIGVRSORDNDYECRNMLKLLKQOORVLLIGISDVSGWGAFLKNSVSKHYL 778  
QY 639 SEYGEIISQDEADRRGVYDKWCSPLFNLNDPVVDAFRKGNKIRFANHSVNPNCYAV 698  
DB 779 GEYGEIISHKPADRKGLTYDRNCSFLFNLNDQVLDAYRKDKLKFANHSPEBNCYAK 838  
QY 699 V 699  
DB 839 V 839

## RESULT 5

152060  
protein MEDA [imported] - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004  
C.Accession: T52060  
R.Groschiklaus, U.; Viejle-Calzada, J.P.; Hoepfner, M.A.; Gagliano, W.B.  
Science 280, 446-450, 1998  
A.Title: Maternal control of embryogenesis by MEDA, a polycomb group gene in Arabidop  
A.Reference number: Z25927; MUID:98212068; PMID:9545225  
A.Accession: T52060

A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-689 <GEO>  
 A/Cross-references: UNIPROT:O65312, UNIPARC:UPI000003C2BE, EMBL:AF060465, PIDD:AA039446,  
 C/Genetics:  
 A/Genes: MEA

Query Match 16.1%; Score 612; DB 2; Length 689;  
 Best local similarity 25.4%; Pred. No. 1.5e-29;  
 Matches 162; Conservative 100; Mismatches 197; Indels 180; Gaps 21;

```

QY 120 FMVDETVLNIPIYMGDEVLDQDGTFI---BELIKNYDKVHGDECGFINDIEFVELV 175
DB 114 FLDEVDPLPSVKPIYKELPRSTIYVFTKSSQLMAESD-SVIGKRIYIYNGE----- 166
QY 176 NALGQYNDDDDDDDDDPPE-REEKQKLEDRDDEKSRPRKPSDSIFPAISMPFDK 234
DB 167 -ALBISSEBDEBDEBEIKKEKCEPSED-----VDRFIWVGQDY 208
QY 235 G-----TABELKEKYKELTEQQLPGALPPECTPNDGPNKASVQREQL 278
DB 209 GLDILVVRALAKTIEVDVSDIERNYELKLN-----DGTAGEA--SDLTIS 253
QY 279 HSFTLTF-----CRRCFKYDCFLAPFHATPTYKKTETAL--DNKPCGQCQY 327
DB 254 KITTAFODPADRRRCRCMIYDCMHBEKY-EPESRSESDKSLFEDEDRQPCSEHCYIK 312
QY 328 LEGAKEPAAALTAELIKTPPRKPGRRGRRLPNSSRPSTPTINVLSEKDTSDREAGTE 387
DB 313 VRSVTE-----ADHV-----MDNDS----- 328
QY 368 TGEENNDKEBEKDEKSSSSSEANSRCQPIKMKNIIEPPENVEWSGAEMFRVLIGTY 447
DB 329 -----LSNKIVSDPNNMTPTPEKOLYLKGIELF 358
QY 448 YDNFCALAR--LIGTKCRQVTEFVKSSSI-----IAPAEVDV 486
DB 359 GNSCDVALNITLKGKTCLEIYVNRBDQCTMSIDLNKTQGRHNQVTKYSRKSRSRY- 417
QY 487 TPRKRRKRLMAHCKRIQLKDGSSNHYVNYPCDHPRQPCSSGCPVCAONFCEKF 546
DB 418 ---KKRKLKRYAYPRPALKKTTSGEAFYRHYTPCT-CKSKCQOQCCLTHENCERY 472
QY 547 COCSSECONRFPGCRCCK-AQCNTKQCPCYLAVERCDPLC-----ITCGAADMHSK-NVS 600
DB 473 CGCSKDCNNRFGSCAICGCTNNRQCPCPAANRECDPLCRSCPLSCGGGTGGERPVQIQ 532
QY 601 CNKCSIQRGSKKHLIAPSDVAGWGIPIKDPVQKNFISEYCGEIIISQDEADRGRKYDK 660
DB 533 CKNMQFLQTNKKIILIGKSDVHGWGAFIWDLSKKIHEYLGEYGEIITHDENRGRIBDR 592
QY 661 YMCSEFLNNDPVVDATREKNIKIRPANSVNPNCYAKV 699
DB 593 IGSSTYLFILNDQLIDARKKNEFKFLNHSANPCYAKL 631

```

## RESULT 6

721436  
 Hypothetical protein R06A4.7 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: T21436; T23953  
 R/Barlow, K.  
 submitted to the EMBL Data Library, November 1996  
 A/Reference number: Z19421  
 A/Accession: T21436  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-775 <WIL>  
 A/Cross-references: UNIPARC:UPI000017BAD; EMBL:Z81515; PIDD:CA04199.1; GSPDB:GN00020;  
 A/Experimental source: clone F26H11  
 R/Barlow, K.  
 submitted to the EMBL Data Library, November 1996

A/Reference number: Z19822

A/Accession: T23953  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-775 <W12>  
 A/Cross-references: UNIPARC:UPI000017BAD; EMBL:Z83120; PIDD:CA05589.1; GSPDB:GN00020;  
 A/Experimental source: clone R06A4

A/Map position: 2  
 A/Introns: 223/1; 414/3; 464/2; 537/1; 601/2; 655/1; 679/1; 749/2

Query Match 13.6%; Score 519; DB 2; Length 775;  
 Best local similarity 24.8%; Pred. No. 7e-24;  
 Matches 162; Conservative 84; Mismatches 226; Indels 170; Gaps 25;

```

QY 106 SVPIYMSPILOQFMVDETVLNIPIYMGDEVLDQDGTFI---BELIKNYDKVHG-DEECG 164
DB 170 SPPIY-WPIEVSVAIPDQRLTHMYPFDGI--DGNIEHLIDMPPGIGHFSDNWS 226
QY 165 FINDERVELVNALGQYNDDDDDDDGDPREBEKQKLEDRDDEKSRPRKPSDKIF 224
DB 227 YVNDWILYKLCRAA-----LKYQGS-----PVFVY 252
QY 225 EATSMPPDKTAELEKRYKELTEQQL--PGALP-----ECTPNDGPNKASV 272
DB 253 YTLIRLWPNKNSQREFSSAPFVLCENPAKGFDPSSLEPWKTKIAGAONLNPFT---- 308
QY 273 QREQLSHFTLFCRCRCFYKDCFLAPFA-----TPNTYKRNTEAL 315
DB 309 -----CYACLAITYCAIHGFKAEPIEFPGHFNAMLPEN-----NPNEN-- 348
QY 316 DNKPCGQCQYHLEGATEPAALATAE-----RIKTPPKRPGRRGRRLP 359
DB 349 DGMKSGNCKRSYV-MKEVSEVLVPDSEELQKEVKIYFMKSIYAKNPIDGALIVYV 407
QY 360 NNS--SRPSTPTINVLSEKDTSDREAGTERGSENNKKEBEKDEKSSSEANSRCQTP 417
DB 408 FNTYIVRFCEV-----KKYVEDDESCKTRSCDATHLM 443
QY 418 IKMKPNTE-----PPENVEWSGAEMFRVLIGTYNDNCALATRLIGTKCRQVTEFR 470
DB 444 MSMAENVSARKLNGQDEN-----RLSTIKDRNNNR-----NLSQEKAKRK 486
QY 471 VKESIAPAPAEVDVTPPRKKRRLMAHCKRIQLKDGSSNHYVNYPCCHPRQPC 530
DB 487 LRHDSLRIGLRDGLDAEKLIRDDMDRDSQNSSEKVRM-----TAVTPITACRH-AGPC 539
QY 531 DSSCP-CYIAQN-FCRKCQCSSGCONRFPGCRCCKA-QCNTKQCPCYLAVERCDPLCT 587
DB 540 NATAENCACRENGVCSYCKCDINCSQRPFCNCAGQCTYKACOCCTPAANMECNPTCNM 599
QY 588 CGAADHMSKVVSCNCSIQGSKKHLIAPSDVAGWGIPIKDPVQKNFISEYCGEIIIS 647
DB 600 C-KCDALISNIKKCNPFQMTMIQKRYCGSKLAGNGLFLLEPAKEKDFITEYGERIS 658
QY 648 QDEADRGRKYDKTNGSEFLNNDPVVDATREKNIKIRPANSVNPNCYAKV 698
DB 659 DDEARRGAIYDRYQCSYIFNIEGGAIDSYKIGLTAAPAHDSKNPTCYAR 710

```

## RESULT 7

741282  
 probable transcription silencing protein - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T41282  
 R/Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, April 1999  
 A/Reference number: Z21984  
 A/Accession: T41282  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA



A:Residues: 1-3968 <TKA>  
A:Cross-references: UNIPROT:Q03164; UNIPARC:UPI000017C3FD; GB:L04284; NID:G184393; PIDN:  
A:Note: sequence extracted from NCBI backbone (NCBI:117729)  
R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canani, O.; Cimino, G.; Croce, C.M.; Car  
Cell 71, 701-708, 1992  
A:Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 g  
A:Reference number: A44264; MUID:93046668; PMID:1432625  
A:Accession: A44264  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 63-316, 'GLINSELEK', 327, 'Q', 329, 'VR', 332, 'DKGNPP', 340, 'T', 342, 'EDKTVROSPP  
546, 'LQIESRSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GUL>  
A:Cross-references: UNIPARC:UPI000017C3FE; GB:L04731; NID:9339921  
A:Note: sequence extracted from NCBI backbone (NCBI:117779)  
R:Diabali, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.  
Nature Genet. 2, 113-118, 1992  
A:Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in acut  
A:Reference number: 158112; MUID:93265134; PMID:1303259  
A:Accession: 158112  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328 <DJA>  
A:Cross-references: UNIPARC:UPI000016B134; GB:L01986; NID:9307522; PIDN:AAA92511.1; PI  
R:Marshall, R.; Grell, J.; Lochner, K.; Nilsson, I.; Siegler, G.; Zweckbronner, I.; Bec  
Br. J. Haematol. 90, 308-320, 1995  
A:Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in the  
A:Reference number: 137165; MUID:95315013; PMID:7794749  
A:Accession: 137165  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1212-1603, 'GTE', <MAR>  
A:Cross-references: UNIPARC:UPI000016A531; EMBL:X83604; NID:9897757; PIDN:CAA58584.1; PI  
A:Note: submitted to the EMBL/GenBank/DBJ databases by R. Marschalek, 20 December 1994  
R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canani, O.; Saito, H.;  
Cancer Res. 54, 2327-2330, 1994  
A:Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved i  
A:Reference number: 138485  
A:Accession: 138485  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1251-1486, 'G', 1488-1538 <RS>  
A:Cross-references: UNIPARC:UPI000016A07B; EMBL:U04737; NID:9451554; PID:9451555  
C:Genetics:  
A:Gene: GDB:MLT; HTX; ALL-1; HRX  
A:Cross-references: GDB:128819; OMIM:159555  
A:Map position: 11q23-11q23  
A:Introns: 1338/1, 1362/3, 1406/3, 1444/3, 1493/3, 1525/3, 1566/1  
A:Note: the list of introns is incomplete  
C:Superfamily: histone methyltransferase, trithorax protein type  
C:Keywords: acute lymphoblastic leukemia, proto-oncogene, zinc finger  
F:144-1456/Region: zinc finger CCHC motif  
F:1479-1506/Region: zinc finger CCHC motif  
F:1527-1556/Region: zinc finger CCHC motif  
F:1569-1596/Region: zinc finger CCHC motif  
F:1873-1900/Region: zinc finger CCHC motif  
F:1933-1955/Region: zinc finger CCHC motif

Query Match 6.0%; Score 228; DB 2; Length 3968;  
Best Local Similarity 18.4%; Pred. No. 1.6e-05;  
Matches 116; Conservative 80; Mismatches 190; Indels 246; Gaps 23;

DB 207 RDKKSRPRKPPDKIFEAISMPDPDGTAE----- 238  
DB 3392 QDQPALP-----SGWFPQIGTSPSTAITAASSICVLPSTQTGITAAS 3440  
QY 239 ---ELKERYK-----ELTEQQLFGALPPPC---TPNIDGNKSVORESL 278  
DB 3441 PSGEADHYQLOHVNQLASTKGTGHSQRDLDSAGSPQVSNFTQTVADAPNSKGLQONRAL 3500  
QY 279 HSFHLLPGRCKKQDCLFHPFATPNTYKRNKTEALDNKPCGPGCYOHLBSAKKFAAL 338  
DB 3501 SS-----AVQASP-----TSQGSFSSPSGGR-----S 3524

QY 339 TAERIKTPPKRPGRRGRGLPNNSSRPSTPTINVLASKYDS---DREACTGTGGENDK 395  
DB 3525 ASPVPPTPKPKPTKQFQLPDKNGKKKSHLTSSSEAHIPDQETSLTSGTTPG 3584  
QY 396 EREKDETSSTSSANRCCQTP---IQMKNE-----PENNY----- 431  
DB 3585 AEAQDUTASLEOSSQKCCQACQAVLBEVQVTQNPANBQSBKTYVBESNSP 3644  
QY 432 ---WSGARAS-----MFRV----- 442  
DB 3645 LMLMLQGEQRKSEITEKPKKGLVFETSSDDGFOICABSIEDAKSLTDKVGARSNAR 3704  
QY 443 -----LIGTYNDNC-ALKRLGTCTGROYEFKRVKSSIIAPAPEDVTP 488  
DB 3705 LKQLSPAGVNGLRMLGTHDAVFLIQLSGAKKCN-YKRFK-----PEANEP 3755  
QY 489 PRKKGRHRLMAHRCRIQLKQSSHHVNYQPCDPRPCDSCPCVLAQNCFRCQ 548  
DB 3756 PLNPHGSARA-EVHLRK-----SAPDMRFLASKH-RQP-----PENPNDBEEBVO 3801  
QY 549 CSSECONRPGCRCKAOCNTKQCPCYLAVEBCPDLCITCGAADHMDSKNVSCKNCSIOR 608  
DB 3802 LKS-----ARRATGMDLPMRPFH-----LKK 3824  
QY 609 GSKTHLLADSDVAGWGIPTKDPVQKNPFTSEYGEIISQDEADRGRKYD-KTMCSPLP 667  
DB 3825 TSKEAVGVYRSPHGRGLFCRKNIDAGEWYIEVAGNIRISIQDKREKYYDSKIGCYMF 3884  
QY 668 NLNMDFYVDAITRKGNKIRPANHSVNPACYAKV 699  
DB 3885 RIDDSVVDATMGNRARFLTNHSCFPCYSRV 3916

RESULT 10  
571490  
ash1 protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 09-Jul-2004  
C:Accession: S71490; S78442  
R:Trippola, N.; LaJeunesse, D.; Glidea, J.; Shearn, A.  
Genetics 143, 913-928, 1996  
A:Title: The Drosophila ash1 gene product, which is localized at specific sites on poly  
A:Reference number: S71490; MUID:96363924; PMID:875238  
A:Accession: S71490  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-2144 <TRI>  
A:Cross-references: UNIPROT:Q24189; UNIPROT:Q9VM15; UNIPARC:UPI000017BE7A; EMBL:U49439  
R:Trippola, N.; Shearn, A.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: S78442  
A:Accession: S78442  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-119, 'V', 121-611, 'X', 613-2144 <TR>  
A:Cross-references: UNIPARC:UPI000017BE7B; EMBL:U49439  
C:Genetics:  
A:Gene: ash1  
A:Introns: 1592/3; 2083/1

Query Match 5.9%; Score 226.5; DB 2; Length 2144;  
Best Local Similarity 21.3%; Pred. No. 1e-05;  
Matches 124; Conservative 64; Mismatches 169; Indels 205; Gaps 26;

QY 241 KKRYKLTBOOLPAL---PRECTPNTDGPVAKSVOREGSLHSFHT-----LFCRRCF 290  
DB 889 KKKRKSIVBOQTAVIVIEHEBEPD---DDEFLQIRKTRSSNNVNVQAAPRPDCCR-- 944  
QY 291 KYDCFLHPFATPNTYKRNKTEALDNKPCGPGCYOHLBSAKKFAALTAERIKTPPKRP 350  
DB 945 -----VPQAGEARETFFVARTNOK-APR-----LSVVALERLGRPQCPA 981  
QY 351 GGRRRGRLLPNNSSR-----DSTPTINVL-----ESKQTDSDREAGRTGGENNDKE 396

```

Db 982 RGRPRGRPKRNEQBAAPQPPKSEBPIRPAKKRGRQKQPVLEBPPTPPQKKKTM 1041
QY 337 EBEKDELTSSSEANSRCOTPIKMKPNT-----PPR---NVMSG 434
Db 1042 EPNITLPGIDPNTNPSCKIRLKRKNLEAGTQPKKRPVQPVTEBIPPEIPVSGEID 1101
QY 435 AEASMFV-LIGTYDNC-----IARLTQTKCR 464
Db 1102 AEAAKRVDSLPTEHDPAPASHNPQPDYASCESEEDKASTSLRKLKSKVKTTLVA 1161
QY 465 QYVEFRVSSIIAP-----PAEDVTPPKKKRKR-----LWMA 501
Db 1162 GLSHMYQSLMPRPAAKNNKRGLEQVGPASLPPPYCEKYLARTEMDFELPYDIWA 1221
QY 502 HC-----RQLQKDGSSNHY-----NYQCDHPROPCCSSCPVIAQ 540
Db 1222 YTNSTLPTRNVVPSWNYRKR-----TNVYASVRLAGFDHP-----TCNC--K 1265
QY 541 NPKRFPCCSSBCCNRFPGCRCKAQNTKQPCVLAVERCDPDLCTLGAADHMSKXVS 600
Db 1266 NQGR--SCLNCLAR-----MVTESPS--NCPAGE-----K 1295
QY 601 CKNGSIQSGSKHLLAP-----SDVAMGIFIKDPVQKNFISRCGELISODE-ADR 653
Db 1296 CRNQLQGH-----VAGVERFMADKMGVRTLPLAKGYILIEYGEVVTETKRPQR 1350
QY 654 RGVYDKTMCSPFLFNLDPVVDATRKGNKIRPANHSVNPNC 695
Db 1351 MASITLNDTHHYCLHLDGLVLDGQRMGSDCFVHNSCEPNC 1392

```

## RESULT 11

```

A48205
A11-1 protein +GTR form - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 22-Jun-2003
C:Accession: A48205; B48205
R:Ma, Q.; Alder, H.; Nelson, K.K.; Chatterjee, D.; Gu, Y.; Nakamura, T.; Canaan, E.; et al.
Proc. Natl. Acad. Sci. U.S.A. 90, 6350-6354, 1993
A:Title: Analysis of the murine A11-1 gene reveals conserved domains with human ALL-1 at
A:Reference number: A48205; MUID:93317679; PMID:8327517
A:Accession: A48205
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3869 <HAR>
A:Cross-references: UNIPARC:UPI000017C621; GB:L17069
A:Accession: B48205
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1505,1509-3869 <HAR>
A:Cross-references: UNIPARC:UPI000017C622; GB:L17069
C:Genetics:
A:Gene: A11-1
C:Superfamily: histone methyltransferase, trithorax protein type
C:Keywords: alternative splicing; zinc finger

```

```

Query Match 5.9%; Score 226; DB 2; Length 3869;
Best Local Similarity 19.9%; Pred. No. 2e-05;
Matches 124; Conservative 78; Mismatches 193; Indels 228; Gaps 25;

QY 207 RDKESRPRKFPBDKIFEAISMPDCKTAA----- 238
Db 3393 QDQVVALPPS-----SGMFPOLGTSQTPSAAMAATASSICVLPSSTQTAGMTAAS 3341
QY 239 ---ELKEKYK-----ELTEQQ-----LPGALPPECTPNDGPNASVOREQ 276
Db 3342 PRGAHEHYKQQRNGQLAGKTGTLTSQRDDPSAPETQPSITQTAKAANGVSLBNK 3401
QY 277 SLHSFHTLFCRCRCRYDCFLHPFATPNTYRKKTETALDNKPCGPOCYHLBGAKEPAA 336
Db 3402 TLPSAKP-----ASSASPGSSPSGQSSSSSV- GPT-----KP 3435

```

```

QY 337 ALTAERIKTPPRKPGRR-----RGLPNNSRPS-----TPTINVLSEKOT- 378
Db 3436 KPKAKRILPLDCKSVKHKVSHLRTSSBAHPRDTPAPQPSYTRTPRAN-RQQODAA 3494
QY 379 ---DSRBAETETJ-----GENDKBEERK-----DETSSSEANSRCQTPY 418
Db 3495 GVEPQKCECGPAGVAALEPVQATQNPANQENAPKAMBEESGSSPLMLMLQOBQ 3554
QY 419 KMKPNI---EPPEVNSGAAISMFRV----- 442
Db 3555 KRKESITRRKKPKGLVFEISSDDGFCASIBDAMKSLTDKVOBARSNALIKQLSPAGV 3614
QY 443 ---LIGTYVNF-C-AIRALIGTKCRQVYFRVSESSIIAPAPAEVDTPPKKKRKR 497
Db 3615 NGLRNLGILHDAVFLLEQLGAKGCRN-YFRFRK-----PERANPEPLPHGSAR 3665
QY 498 LMAACRKLQKDGSSNHYNYQPCDHPROPCCSSCPVIAQNECFKCCSSBCCNRF 557
Db 3666 A-EVHLRK-----SARDMNPFLASKH-RQP-----PEYNPDBEERVOLKS----- 3705
QY 558 PGCRCQAQNTKQPCVLAVERCDPDLCTLGAADHMSKXVSCNCSIQSGSKHLLA 617
Db 3706 ---ARRATSMDLPMRPFH-----LKTITSKXAVGVY 3734
QY 618 PSDVAMGIFIKDPVQKNFISRCGELISQDBADRGKYVD-KYMGSPFLFNLDPVVD 676
Db 3735 RSPHGRQLPCKRNIDGEMVETAYGAVTNSIQTDKREKTYDSKIGCYMFRIDSEYVD 3794
QY 677 ATRKGNKIRPANHSVNPNCYAKV 699
Db 3795 ATTHGNARFINHSCPEPCYSRV 3817

```

## RESULT 12

```

H88444
protein C26B6.12 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: H88444
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biol.
A:Reference number: A75000; MUID:9065613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_e1
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; ar
A:Accession: H88444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1802 <STO>
A:Cross-references: UNIPARC:UPI000017A5C6; GB:chr_III; PIDN:AAA21166.1; PID:G532813; G5
C:Genetics:
A:Gene: C26B6.12
A:Map position: 3

```

```

Query Match 5.8%; Score 220.5; DB 2; Length 1802;
Best Local Similarity 19.4%; Pred. No. 1.9e-05;
Matches 162; Conservative 93; Mismatches 269; Indels 311; Gaps 30;

QY 105 ASVPINTSWSPLOQNFVWDETVLHNPYMGDEVLDGTFIEELIKNYDGKVGADRECG 164
Db 989 AQVEVEYDY-PLKSHSESHDR--HSLBMDVYV-SSDG-----ETVSN----- 1027
QY 165 FINDEIVELVNALGQYNDDDDDDDGDDPBEERBEKQDLR-----DH 206
Db 1028 -----VEKLECM-----EKKKQDLERIALATPTIVYKCKKRWMD 1063
QY 207 RDKESRPRKFPBDKIFEAISMPDCKTAAELKEKYKELTEQOLGALPPECTPNDG 266
Db 1064 LSRVARDIRQOQIRQCFALDERKLAKALADESKR-KEREKARQBAEKSPNHLIADM 1122
QY 267 PNAKSVOREQSLHSFHTLFCRCRCRYDCFLHPFATPNTYRKKTETALD-----NKPC 320
Db 1123 MPQGTLLVNNQSPASSSGFTRK-QKTRPKSHPKQHHNHAKASVSTRVHSSSTRSSSV 1181

```

```
QY 321 GPCCYCHLEGAKEFAALTAERIK-----TPKRRGRRRGLPNNNSRP-----STPT 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1182 APTGQRVSTSSSSSSAATSRVSEDSDSSTGEGVQRRTSYLSNDKRRBRASFSSTS 1241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 INVLESDTDSDRAGTGTGENDKEBEKKDETS-----SSSANSRCOT----- 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1242 IQSGPERQDVSSSSSTSSSSSTSMKQETADKSRKRLIMSDDSGSTGSTATSVS 1301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 417 -----PIKMKPIEP-----PENVMSGAAASFRVLIG 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1302 SRGSLSEPPQDEKTPGPPKKSGQDPTSERVSKLEGERRPLSEVETSGP----- 11G 1354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 -----TYDNFCAIARLIGTKCR-----QYEFRRV- 471
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1355 DSSYLPYKTYHWEKAGIIEMVLPAHSIRAHYHPTTEHCYFGIDDPQPKIQIFDHSPC 1414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 -----KESSTIAPD-----AR-----DVTPPRKKRKRRLMAHCRKIQLKKG 512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1415 KSEPGSEPLKITPAPWGFIDNVASTGPLIYMDVVTAPVTQVKOK-----PRKQVFEKD- 1468
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 513 SSNNVYVYQPCDHRRQPCDSSCPVIAQNFCEKQCQSSCONRPPG-CR----- 561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1469 -PYETIEPPPTKRPAP-----PRFKCTTPRPSSEKKKIIQDCEDLPLEDMQY 1517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 562 CKAQCNTPKQCPCYLA-----VRECDPDLCLT----- 587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1518 LRALINMQSEVSKADBLPWKKMLTFKMKLSEBPLALNPTRSKGLPDAFYDEBLDG 1577
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 588 -----CGAADHDSKNVSCNCSIORGS----- 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1578 VIPVAAGSRAPPEKMTMKQKRSIVRRPDNESHPTAIFSRDEDTAIRHQHASKDWMLL 1637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 611 -----KKHLLAPSDVAGMGIPLKDPVQKNEPTISEYCEBII 646
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1638 QRLLITSLIGDANNDPFTKINQKFKKMIKPARSRHIGHLVAMEIADENIVETIGQTI 1697
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 647 SQDEADRKKYVDK--YWCSEPLFNANDPVVDATRKNGKIRFANHSVNPNCYAKV 699
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1698 RSLVAEBERKAYERRGIGSSYLFRIIDLHVTDATRGPAFPIHNSCGPNCYAKV 1752
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 13

```
E96795
Unknown protein F28016.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: E96795
R:Title: Logis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
  Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
  Hansen, N.F.; Hughes, B.; Hulcar, L.
  Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
  C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzilli,
  Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
  A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
  ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
  A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
  A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96795
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <STO>
A:Cross-references: UNIPROT:Q9SRE2; UNIPARC:UPI00000A4C99; GB:AB005173; NID:96143888; PI
  C:Genetic:
  A:Gene: F28016.8
  A:Map position: 1
```

```
Query Match 5.8%; Score 219; DB 2; Length 528;
Best Local Similarity 27.2%; Pred. No. 6.3e-06;
Matches 50; Conservative 27; Mismatches 75; Indels 32; Gaps 4;
```

516 HVY--NYQPCDHPROPDSSCPVIAQNFCEKQCQSSCONRPPGCKAQCNTKQCP 573

```
Db 19 HITQNPFSYKHKQKQBEDISICEKPFEDPDPSACGERCLNVI-----TWTECTPGYCPC 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 YLAVRECDPDLCTCGAADHDSKNVSCNCSIORGSKKHLLAPSDVAGMGIPLKDPVQ 633
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -----GVYCKNQKQKCEYATKRLIKCEGRGMGLVALLERIK 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 634 KNEFISEYCEIISQDEADRKKYVDK--MCSFLEFNANDPVVDATRKNGKIRFANHSV 691
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 AGQFIMEYCEGVISWEAKRRAQTYETHGVKDAYIISLNSBAIDATKGSILARFIHSC 170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 692 NPKC 695
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 RPNC 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 14

```
T16601
Hypothetical protein K09F5.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16601
R:Title: Bentley, D.
  submitted to the EMBL Data Library, September 1995
A:Description: The sequence of C. elegans cosmid K09F5.
A:Reference number: Z18544
A:Accession: T16601
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-408 <BRN>
A:Cross-references: UNIPROT:Q21404; UNIPARC:UPI000007BE66; EMBL:U37430; NID:g1019963; P
  A:Experimental source: strain Bristol N2; clone K09F5
A:Genetic:
A:Gene: CESP:K09F5.5
A:Map position: X
A:Introns: 25/3; 56/2; 220/2; 350/1; 370/3
```

Query Match 5.5%; Score 210; DB 2; Length 408;

Best Local Similarity 29.5%; Pred. No. 1.7e-05;

Matches 62; Conservative 22; Mismatches 78; Indels 48; Gaps 7;

```
QY 504 RKTQKKGSSNNHYNYQPCDHPR-OPDSSCPVIAQNFCEKQCQSSCONRPPG--- 559
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 KKTQLYTKKNSVFYNTDPEFFENTRMADYSFLC-----SPKRTGLLT 53
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 560 ---CRKAQCNTKQCPCYLAVRECDPDLCTCGAADHDSKNVSCNCSIORGSKKHLL- 614
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 VTSCKCGTDCITTEBCSNFANHRECPR-----GCSNCENQRFKRPQFCG 96
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 615 ---LLAPSDVAGMGIPLKDPVQKNEPTISEYCEIISQDEADRKKY--DKYMGSPLENTL 669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 VETPLTNGI-GHGLRATREIATATGLILERKGAHITTAENHKVRYKQDGIKHSYSPEV 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 NNDPVVDATRKNGKIRFANHSVNPNCYAKV 699
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 GRNYYVDPTRKGNARFIHNSCNALYKV 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 15

```
A88445
Protein C26E6.10 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: A88445
R:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
  Science 282, 2012-2018, 1998
  Anonymous, The C. elegans Sequencing Consortium.
  A:Reference number: A75000; MUID:99069613; PMID:9851916
  A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.ganger.ac.uk/Projects/C_e
  A:Note: published extra appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a
  A:Accession: A88445
  A:Status: preliminary
  A:Molecule type: DNA
```

A:Residues: 1-739 <STO>  
A:Cross-references: UNIPROT:Q18221; UNIPARC:UPI000007A0F6; GB:chr\_III; PIDN:AAA21163.1;  
C:Genetics:  
A:Gene: C26B6.10  
A:Map position: 3

Query Match 5.5%; Score 208.5; DB 2; Length 739;

Best Local Similarity 20.6%; Pred. No. 3.9e-05;  
Matches 119; Conservative 59; Mismatches 184; Indels 215; Gaps 20;

```
QY 301 ATPNTYKRNKTEALDNKPCGPOCYOHLGAKFPAAALTAERIKTPPKPGGRRGRRLPN 360
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 150 STPGEVQRKTSVLSNDK-----RRRASFSSTSIQSSPERQ--RDVSSSR 194
QY 361 NSSRPSTPTINVLKSDTD-----SDREAGTETGGENNDKE 396
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 195 TSSSSSTSEMQRBTADKSRKRLIMSDBSSTGSTATSVVSRQSSLEPQDEKTDGE 254
QY 397 EEKQDETSSSEANRCQTPIMKKNIPPEP-----VEMSGA----- 435
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 255 PKKKSQTFISERVSKIRG--EERLPPEVETSGPTIGDSSYLPYKIVHMERAGIIEKN 312
QY 436 -EASMERVLIGTYDNFCAIARLIGTKCR---QVYEFV-----KESIIAPAP--- 481
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 313 LPANSIR---AHENHPFTEHCYFGIDDPQPKIQIFDHSCKSERGSEPLKITPAPWGP 369
QY 482 ----AE-----DVTTPRKKKKRRLMAHCRKIQLKDGSSNHVINYQPCDHPGPC 530
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 370 IDNVAETGPLIMDVTAETVQKKQK---PRKQVFED--PYEYEPPTKRPAP- 421
QY 531 DSSCPVIAQNCEKPCQSSCQNRFPQ-CR-----CKAQNTKQCPCTLA--- 576
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 422 -----PRFKTTPKRSBEKKKLIIGDCEDLPDLEDQWYLRALNEMQSEVKSADKL 472
QY 577 -----VRECDPDLCLT-----CGAADHWDSKN 598
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 473 PKKKMLTFKEMLRSEBPLRLNPIRSKGLPDAPFYDEBLDGVIPVAGCSRARPYEKWT 532
QY 599 VSCNCSIORGS----- 610
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 533 MKQKRSLYRRPNESHPTAIFSERDETAIRHOLASKDRLQRRLLTSLGDANDFFKLI 592
QY 611 -----KHLTLAPSDVAGMGIFIKDPVQKNEFISFGELISODEADRGKVVYDK--YM 662
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 593 NQTKRKQKTKPARSRIRHGMGLYAMESIAPDEMTIVETIGQTRSLVAERERKAYERRGTG 652
QY 663 CSPLFNANDFVVDATRKGNKIRFANHSTVNPCTYAKV 699
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 653 S6YLFRIDLHVIDATKGNFARFINHSGQPCYAKV 689
```

Search completed: August 11, 2006, 22:29:11  
Job time : 32.573 sec



PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:17:04 ; Search time 187.206 Seconds  
(without alignments)  
3453.870 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699  
Perfect score: 3808  
Sequence: 1 MGQCKSKSEKPGVCKMKRKYK.....KMKTRFANHSVNPICVAKV 699

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 92501592 residues  
Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	746	1 EHZ2_HUMAN	Q15910 homo sapien
2	3803	99.9	746	2 Q4R381_MACFA	Q4R381 macaca fasc
3	3787.5	99.5	751	2 Q96F16_HUMAN	Q96F16 homo sapien
4	3756	98.6	746	2 Q99L74_MOUSE	Q99L74 mus musculu
5	3756	98.6	779	2 Q571L5_MOUSE	Q571L5 mus musculu
6	3750	98.5	746	1 EHZ2_MOUSE	Q61188 mus musculu
7	3747	98.4	746	2 Q3T2H6_MOUSE	Q3T2H6 mus musculu
8	3679	96.6	742	2 Q6AKH7_MOUSE	Q6AKH7 mus musculu
9	3566	93.6	748	2 Q5XHM0_XENTLA	Q5XHM0 xenopus lae
10	3550	93.2	748	2 Q98SM3_XENTLA	Q98SM3 xenopus lae
11	3515	92.3	748	2 Q4V863_XENTLA	Q4V863 xenopus lae
12	3455.5	90.7	695	2 Q4R780_MACFA	Q4R780 macaca fasc
13	3388	88.0	664	2 Q75MP9_HUMAN	Q75MP9 homo sapien
14	3367.5	88.4	669	2 Q75M00_HUMAN	Q75M00 homo sapien
15	3164.5	83.1	759	2 Q90WP4_TETNG	Q90WP4 tetracodon n
16	3138	82.4	760	2 Q5TKR5_ORYLA	Q5TKR5 oryzias lat
17	2389.5	62.7	747	1 EHZ1_MOUSE	P70351 mus musculu
18	2389.5	62.7	750	2 Q3U3V5_MOUSE	Q3U3V5 mus musculu
19	2389.5	62.7	750	2 Q922L1_MOUSE	Q922L1 mus musculu
20	2386.5	62.7	747	1 EHZ1_HUMAN	Q92800 homo sapien
21	2386.5	62.7	747	2 Q53XP3_HUMAN	Q53XP3 homo sapien
22	2381.5	62.5	747	2 Q5RDS6_PONPY	Q5RDS6 pongo pygma
23	2278.5	59.8	683	2 Q3UW02_MOUSE	Q3UW02 mus musculu
24	2212.5	58.1	782	2 Q48976_TETNG	Q48976 tetracodon n
25	2082	54.7	508	2 Q504B3_BRARE	Q504B3 brachydanto
26	2048	53.8	391	2 Q3U575_MOUSE	Q3U575 mus musculu
27	1959.5	51.5	760	1 E2_DROME	P41224 drosophila
28	1945.5	51.1	742	2 Q7PTY9_ANGCA	Q7PTY9 anopheles g
29	1908	50.1	749	2 Q2U2J3_DROPS	Q2U2J3 drosophila
30	1768	46.4	376	2 Q59H64_HUMAN	Q59H64 homo sapien
31	1631.5	42.8	579	2 Q8BR85_MOUSE	Q8BR85 mus musculu

32	1631.5	42.8	582	2 Q3TPR1_MOUSE	Q3TPR1 mus musculu
33	1515	39.8	286	2 Q6R125_HUMAN	Q6R125 homo sapien
34	1290.5	33.9	354	2 Q4T531_TETNG	Q4T531 tetracodon n
35	1261.5	33.1	346	2 Q4T5B4_TETNG	Q4T5B4 tetracodon n
36	1163.5	30.6	231	2 Q8C215_MOUSE	Q8C215 mus musculu
37	1141.5	30.0	312	2 Q4SVV6_TETNG	Q4SVV6 tetracodon n
38	1136	29.8	395	2 Q4TFJ6_TETNG	Q4TFJ6 tetracodon n
39	1105	29.0	247	2 Q6GNT4_XENTLA	Q6GNT4 xenopus lae
40	1080	28.4	269	2 Q4TJES_TETNG	Q4TJES tetracodon n
41	1026.5	27.0	267	2 Q4TF58_TETNG	Q4TF58 tetracodon n
42	918	24.1	246	2 Q4T9K7_TETNG	Q4T9K7 tetracodon n
43	791.5	20.8	277	2 Q4RDS8_TETNG	Q4RDS8 tetracodon n
44	727	19.1	814	2 Q76194_PETRY	Q76194 petunia hyb
45	719.5	18.9	856	1 E2AL_ARATH	Q9ZSM8 arabidopsis

## ALIGNMENTS

RESULT 1  
EHZ2\_HUMAN STANDARD; PRT; 746 AA.  
ID EHZ2\_HUMAN Q15910; Q15755; Q92857;  
AC Q15910; Q15755; Q92857;  
DT 15-JUL-1998, integrated into UniprotKB/Swiss-Prot.  
DT 15-JUL-1998, sequence version 2.  
DT 07-FEB-2006, entry version 44.  
DE Enhancer of zeste homolog 2 (EMX-1).  
GN Name=EHZ2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
OC Homo.  
NCBI\_TaxID=9606;  
OK NCBI [1]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=97357309; PubMed=9214638; DOI=10.1093/emboj/16.11.3219;  
RC Tissue=Brain;  
RX MEDLINE=97124843; PubMed=8954776; DOI=10.1006/geno.1996.0588;  
RA Chen H., Rosster C., Antonarakis S.B.; DOI=10.1006/geno.1996.0588;  
RT "Cloning of a human homolog of the Drosophila enhancer of zeste gene (EHZ2) that maps to chromosome 21q22.2.";  
RL Genomics 38:30-37(1996).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=97357309; PubMed=9214638; DOI=10.1093/emboj/16.11.3219;  
RA Laidle G., Wolf A., Dorn R., Reuter G., Nislow C., Lebererger A.,  
RX Popkin D., Pillus L., Jenuwein T.;  
RT "Mammalian homologues of the Polycomb-group gene Enhancer of zeste  
mediate gene silencing in Drosophila heterochromatin and at S.  
cerevisiae telomeres.";  
RL EMBO J. 16:3219-3232(1997).  
[3]  
RN NUCLEOTIDE SEQUENCE OF 134-746.  
RX MEDLINE=96220494; PubMed=8649418;  
RA Hobert O., Jallat B., Ullrich A.;  
RT "Interaction of Vav with ERK-1, a putative transcriptional regulator  
of homeobox gene expression.";  
RL Mol. Cell. Biol. 16:3066-3073(1996).  
[4]  
RN ATRX BINDING.  
RX MEDLINE=98167853; PubMed=9499421; DOI=10.1093/hmg/7.4.679;  
RA Cardoso C., Timmlt S., Villard L., Khrestchatsky M., Fontes M.,  
RL Colleaux U.;  
RT "Specific interaction between the XNP/ATRX-X gene product and the SRT  
domain of the human EHZ2 protein.";  
RL Hum. Mol. Genet. 7:679-684(1998).  
[5]  
RN PHOSPHORYLATION SITE THR-487.  
RX PubMed=15302935; DOI=10.1073/pnas.0404720101;  
RA Beausoleil S.A., Jedrychowski M., Schwartz D., Elias J.B., Villen J.,  
RL Li J., Cohn M.A., Cantley L.C., Gygi S.P.;  
RT "Large-scale characterization of HeLa cell nuclear phosphoproteins.";  
Proc. Natl. Acad. Sci. U.S.A. 101:12130-12135(2004).  
CC 1- FUNCTION: May be involved in the regulation of gene transcription

CC and chromatin structure.  
 CC -1- SUBUNIT: Binds ATPX via the SET domain (Probable).  
 CC -1- SUBCELLULAR LOCATION: Nucleus (Probable).  
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues.  
 CC -1- SIMILARITY: Belongs to the EZ family.  
 CC -1- SIMILARITY: Contains 1 SET domain.  
 -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs license  
 -----  
 DR EMBL: X95653; CA64955.1; -; mRNA.  
 DR EMBL: U61145; AAC51520.1; -; mRNA.  
 DR PIR: G02838; G02838.  
 DR HSSP: O8X225; 1M.9.  
 DR TRANSFAC: T04888; -.  
 DR Ensembl: ENSG00000106462; Homo sapiens.  
 DR HGN: HGNC:3527; EZH2.  
 DR MIM: 601573; gene.  
 DR GO: GO:0003677; P:DNA binding; TAS.  
 DR GO: GO:0006325; P:establishment and/or maintenance of chromatin; TAS.  
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro: IPR001005; MYD\_DNA\_Bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS00280; SET; 1.  
 KM DNA-binding; Nuclear protein; Phosphorylation; Transcription;  
 KM Transcription regulation.  
 FT CHAIN 1 746  
 FT DOMAIN 611 731  
 FT MOTIF 490 495  
 FT COMBINS 523 605  
 FT MOD RES 487 487  
 FT CONFLICT 224 224  
 FT CONFLICT 724 724  
 FT SEQUENCE 746 AA; 85363 MW; 1B5029BB9D509BES CRC64;  
 SQ  
 Query Match 100.0%; Score 3808; DB 1; Length 746;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-214; Indels 0; Gaps 0;  
 Matches 699; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGQTGKSEKGVPCRKRVKSEYMLRLQLKFRADDEVKSPSSNRQKILERTETLNDW 60  
 DB 1 MGQTGKSEKGVPCRKRVKSEYMLRLQLKFRADDEVKSPSSNRQKILERTETLNDW 60  
 QY 61 KORRIQPHILTSVSLRGTRCSVTSDLDPPTQVIPLKTLNAVASVPIWMSPLQONF 120  
 DB 61 KORRIQPHILTSVSLRGTRCSVTSDLDPPTQVIPLKTLNAVASVPIWMSPLQONF 120  
 QY 121 MWEDETVLAHNTPYMGDEVLDQGTETIELIKNYDKVHDCGFTINBEIPELVNAGQ 180  
 DB 121 MWEDETVLAHNTPYMGDEVLDQGTETIELIKNYDKVHDCGFTINBEIPELVNAGQ 180  
 QY 181 YVDDDDDDGDDPEREEREKQDLKDRDDEKSRPRKPPSDKIFALISMPFDKGTAEEL 240  
 DB 181 YVDDDDDDGDDPEREEREKQDLKDRDDEKSRPRKPPSDKIFALISMPFDKGTAEEL 240  
 QY 241 KEKYKELTEQQLPGALPPECTPNIDGPNKSVQREQSILHSFHTLFCRCFKYDCLAPPH 300  
 DB 241 KEKYKELTEQQLPGALPPECTPNIDGPNKSVQREQSILHSFHTLFCRCFKYDCLAPPH 300  
 QY 301 ATPNTYKRNKTETALDNKPCGQCQOHLGAKEPAAALTAERIKTPPRXRGRRGRLEN 360  
 DB 301 ATPNTYKRNKTETALDNKPCGQCQOHLGAKEPAAALTAERIKTPPRXRGRRGRLEN 360  
 QY 361 NSSRSTPTINVLASQDTSRBAETGTGEGNNDKEEEKQDETSSSSANRCQPTIKM 420  
 DB 361 NSSRSTPTINVLASQDTSRBAETGTGEGNNDKEEEKQDETSSSSANRCQPTIKM 420  
 QY 421 KENIEPENVEWGAEMFRVLIGTYDNFCAIALRLIGTKTCROYTEFRVRESSTIAPA 480

DB 421 KENIEPENVEWGAEMFRVLIGTYDNFCAIALRLIGTKTCROYTEFRVRESSTIAPA 480  
 QY 481 PAEDVDTPPRKCRKRLMAAHGKRIQLKDGSSNNHYNTOPCDHPQPCDSCPCYIAQ 540  
 DB 481 PAEDVDTPPRKCRKRLMAAHGKRIQLKDGSSNNHYNTOPCDHPQPCDSCPCYIAQ 540  
 QY 541 NPKCFQCCSSECONRPPGCRCAQCNTKOCPCYLAVERCDPDLCTCGAADHWDNRVS 600  
 DB 541 NPKCFQCCSSECONRPPGCRCAQCNTKOCPCYLAVERCDPDLCTCGAADHWDNRVS 600  
 QY 601 CKKCSIORGSKHLLAPSDVAGWGIFIKDPVQNEFTISEYCGEIIISQDEADRGRKYDK 660  
 DB 601 CKKCSIORGSKHLLAPSDVAGWGIFIKDPVQNEFTISEYCGEIIISQDEADRGRKYDK 660  
 QY 661 YMCSEFLNLDNFVVDATRKGNKIRFANHSVNPVCYAKV 699  
 DB 661 YMCSEFLNLDNFVVDATRKGNKIRFANHSVNPVCYAKV 699  
 -----  
 RESULT 2  
 QAR381\_MACFA PRELIMINARY; PRT; 746 AA.  
 ID QAR381\_MACFA  
 AC QAR381; 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Testis cDNA clone: Qc8A-18821, similar to human enhancer of zeste homolog 2 (Drosophila) (EZH2), transcript variant 2.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopitheciinae; Macaca.  
 OC NCBI\_TaxID=9541;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed:1594441; DOI=10.1093/molbev/ms1187.  
 RA Osada N., Hirata M., Tanuma R., Kusuda J., Hida M., Suzuki Y., Sugano S., Gotohori T., Shen C.-R.J., Wu C.-I., Hashimoto K.; "Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis between Human and Cynomolgus Monkey cDNAs."; Mol. Biol. Evol. 22:1976-1982(2005).  
 RL [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP International consortium for macaque cDNA sequencing and analysis;  
 RG "DNA sequences of macaque genes expressed in brain or testis and its evolutionary implications."; Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs license  
 -----  
 CC EMBL: AB179385; BAB02436.1; -; mRNA.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; P:DNA binding; IEA.  
 DR InterPro: IPR001005; MYD\_DNA\_Bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS00280; SET; 1.  
 SQ SEQUENCE 746 AA; 85313 MW; 169421996E75621 CRC64;  
 Query Match 99.9%; Score 3803; DB 2; Length 746;  
 Best Local Similarity 99.9%; Pred. No. 2.6e-214;  
 Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MGQTGKSEKGVPCRKRVKSEYMLRLQLKFRADDEVKSPSSNRQKILERTETLNDW 60  
 DB 1 MGQTGKSEKGVPCRKRVKSEYMLRLQLKFRADDEVKSPSSNRQKILERTETLNDW 60  
 QY 61 KORRIQPHILTSVSLRGTRCSVTSDLDPPTQVIPLKTLNAVASVPIWMSPLQONF 120

Db 61 KORR1QP4H1LTSVSLGTRBCTVSDLPPTQV1PLKTLNAAVAVPIMWSPLQONF 120  
 QY 121 MWEDEVLANIPYMGDEVLDGGTFTIBELIKNYDGVDHGRGCFINDIIFVELNALGQ 180  
 Db 121 MWEDEVLANIPYMGDEVLDGGTFTIBELIKNYDGVDHGRGCFINDIIFVELNALGQ 180  
 QY 181 YNDDDDDDGDDPBERBERKQDLBDHDDKESRPPKPSDKIFRAISSMPFDKGTAEEL 240  
 Db 181 YNDDDDDDGDDPBERBERKQDLBDHDDKESRPPKPSDKIFRAISSMPFDKGTAEEL 240  
 QY 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCFKYDCFLAPPH 300  
 Db 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCFKYDCFLAPPH 300  
 QY 301 AHPNTYKRTKRTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRRLN 360  
 Db 301 AHPNAYKRNTEFTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRRLN 360  
 QY 361 NSRPSTPTINVLKSDTSDRAGTETGGENNDKEBERKDETSSSSRANRCOTPIK 420  
 Db 361 NSRPSTPTINVLKSDTSDRAGTETGGENNDKEBERKDETSSSSRANRCOTPIK 420  
 QY 421 KENIPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKSIIABA 480  
 Db 421 KENIPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKSIIABA 480  
 QY 481 PABVDVTPPKKKRKRRLMAAHCRKIQLKDGSSNNVYNYQPCDHPROQDSSCPVIAQ 540  
 Db 481 PABVDVTPPKKKRKRRLMAAHCRKIQLKDGSSNNVYNYQPCDHPROQDSSCPVIAQ 540  
 QY 541 NCEKPCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMSKYNV 600  
 Db 541 NCEKPCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMSKYNV 600  
 QY 601 CKNCSIORGSKGHLAPSDVAGWGIPIKDPOVQKNEFISYGBEIIISQDEADRGKRYDK 660  
 Db 601 CKNCSIORGSKGHLAPSDVAGWGIPIKDPOVQKNEFISYGBEIIISQDEADRGKRYDK 660  
 QY 661 YNCSPLFNANDPVVDATRKANKIRPANSVNPNCYAKV 699  
 Db 661 YNCSPLFNANDPVVDATRKANKIRPANSVNPNCYAKV 699  
 RESULT 3  
 ID Q96F16 HUMAN PRELIMINARY; PRT; 751 AA.  
 AC Q96F16;  
 DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-DEC-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 24.  
 DE Enhancer of zeste 2, isoform a.  
 GN Name=EZH2;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;  
 OC Homo.  
 NCBI\_TaxID=9606;  
 OK NCBI\_TaxID=9606;  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=pancreas;  
 RC MEDLINE=22389257; PubMed=12479932; DOI=10.1073/pnas.242603899;  
 RX STRAUSBERG R.L., Feingold B.A., Gronow L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemien C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.T., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richarde S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=pancreas;  
 RG NIH MGC Project;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs license  
 CC -----  
 DR EMBL: BC010858; AAH10858.1; -; mRNA.  
 DR HSSP: Q8X225; 1ML9.  
 DR Ensembl: ENSG00000106462; Homo sapiens.  
 DR GO: 0005634; C:nucleus; IEA.  
 DR GO: 0003677; F:DNA binding; IEA.  
 DR InterPro: IPR01005; MYD\_DNA\_Bd.  
 DR InterPro: IPR01214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS0280; SET; 1.  
 DR SEQUENCE 751 AA; 86018 MW; D865CF02E60E836 CRC64;

Query Match 99.5%; Score 3787.5; DB 2; Length 751;  
 Best Local Similarity 99.1%; Pred. No. 2,1e-213;  
 Matches 698; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 MGQTKSKSEKGPVCKRKKVKSRYMLRQLKFRADRYKSWFSSNRQKILERTETLNDSEW 60  
 Db 1 MGQTKSKSEKGPVCKRKKVKSRYMLRQLKFRADRYKSWFSSNRQKILERTETLNDSEW 60  
 QY 61 KORR1QP4H1LTSVSLGTRBCTVSDLPPTQV1PLKTLNAAVAVPIMWSPLQONF 120  
 Db 61 KORR1QP4H1LTSVSLGTRBCTVSDLPPTQV1PLKTLNAAVAVPIMWSPLQONF 120  
 QY 121 MWEDEVLANIPYMGDEVLDGGTFTIBELIKNYDGVDHGRGCFINDIIFVELNALGQ 180  
 Db 121 MWEDEVLANIPYMGDEVLDGGTFTIBELIKNYDGVDHGRGCFINDIIFVELNALGQ 180  
 QY 181 YNDDDDDDGDDPBERBERKQDLBDHDDKESRPPKPSDKIFRAISSMPFDKGTAEEL 240  
 Db 181 YNDDDDDDGDDPBERBERKQDLBDHDDKESRPPKPSDKIFRAISSMPFDKGTAEEL 240  
 QY 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCFKYDCFLH--- 297  
 Db 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCFKYDCFLH--- 297  
 QY 298 --PFAATPTNTYKRTKRTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRR 355  
 Db 298 --PFAATPTNTYKRTKRTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRR 355  
 QY 301 NYSFATPTNTYKRTKRTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRR 360  
 Db 301 NYSFATPTNTYKRTKRTALDNKPCGQCQYOHLEGAKEPAALTAERIKTPPKRPGRRR 360  
 QY 356 GRLPNSSSRPSTPTINVLKSDTSDRAGTETGGENNDKEBERKDETSSSSRANRCQ 415  
 Db 356 GRLPNSSSRPSTPTINVLKSDTSDRAGTETGGENNDKEBERKDETSSSSRANRCQ 415  
 QY 416 TPRIKKNIPPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKS 475  
 Db 416 TPRIKKNIPPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKS 475  
 QY 421 TPRIKKNIPPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKS 480  
 Db 421 TPRIKKNIPPEPENWVGASAFMFRVLIGTYDNFCAIARLIGTKCRQYERFRVKS 480  
 QY 476 IIAAPADVDVTPPKKKRKRRLMAAHCRKIQLKDGSSNNVYNYQPCDHPROQDSSCP 535  
 Db 476 IIAAPADVDVTPPKKKRKRRLMAAHCRKIQLKDGSSNNVYNYQPCDHPROQDSSCP 535  
 QY 536 CVIAGNFCCKCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMD 595  
 Db 536 CVIAGNFCCKCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMD 595  
 QY 541 CVIAGNFCCKCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMD 600  
 Db 541 CVIAGNFCCKCQCCSSCONRFPGCRCAQCNCTQCPCYLAVRBCDPLCLTCGAADHMD 600

QY 596 SKVNSCKNCSIORGSKHLLAPSDVAGMGIFIKDPVOKNFISEYCGEIIISODEADRG 655  
| | | | |  
DB 601 SKVNSCKNCSIORGSKHLLAPSDVAGMGIFIKDPVOKNFISEYCGEIIISODEADRG 660  
| | | | |  
QY 656 KYVDKMGCSFLPNLNDPVVDATRKGNKIRFANHSVNPCTAKV 699  
| | | | |  
DB 661 KYVDKMGCSFLPNLNDPVVDATRKGNKIRFANHSVNPCTAKV 704  
| | | | |  
RESULT 4  
Q99L74\_MOUSE PRELIMINARY; PRT; 746 AA.  
ID Q99L74\_MOUSE  
AC Q99L74\_MOUSE  
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2001, sequence version 1.  
DT 07-FEB-2006, entry version 28.  
DE Enhancer of zeste homolog 2.  
GN Name=Bzh2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
arose spontaneously;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Datchenko L., Murina N.K., Farmer A.A., Rubin G.M., Hong L.,  
Stephenson M., Soares M.B., Bonaldi M.P., Casavant T.L., Scheetz T.R.,  
Brownstein M.J., Uebachs T.B., Toshiyuki S., Carninci P., Prange C.,  
Rosa S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullanb S.J.,  
Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Rachalski S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Rahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.B.,  
Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
arose spontaneously;  
RA Strausberg R.L.  
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN-CZECH II; TISSUE=Mammary tumor metastasized to lung. Tumor  
arose spontaneously;  
RA Director MGC Project;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs license  
-----  
CC EMBL; BC003772; AA03772.1; -; mRNA.  
DR EMBL; BC016391; AA016391.1; -; mRNA.  
DR HSSP; Q8X225; 1ML3.  
DR Ensembl; ENSMUSG00000029687; Mus musculus.  
DR MGI; MGI:107940; Bzh2.  
DR GO; GO:0045120; Cytonucleus; IDA.  
DR GO; GO:0003682; F:chromatin binding; IDA.  
DR GO; GO:0042054; F:histone methyltransferase activity; IDA.

DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0016571; F:histone methylation; IDA.  
DR GO; GO:0051154; P:negative regulation of striated muscle cell. . .; IDA.  
DR InterPro; IPR001005; Mtb\_DNA\_bd.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00717; SANT; 2.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS50280; SET; 1.  
SQ SEQUENCE 746 AA; 85292 MW; 7442C751E13EA24B CRC64;  
Query Match 98.6%; Score 3756; DB 2; Length 746;  
Best Local Similarity 98.3%; Pred. No. 1.5e-211;  
Matches 687; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MGQTGKSEKGPVCMRKRVKSEYWRLLQRLFRFRADSVKMSFSSNRKILERTETLNQEW 60  
| | | | |  
DB 1 MGQTGKSEKGPVCMRKRVKSEYWRLLQRLFRFRADSVKMTFSSNRKILERTETLNQEW 60  
| | | | |  
QY 61 KORRIQPVHILTSVSLRGTRGCVTSDDLPPTQVIFPLKTLNVAASVPIMYSPLQONF 120  
| | | | |  
DB 61 KORRIQPVHILTSVSLRGTRGCVTSDDLPPTQVIFPLKTLNVAASVPIMYSPLQONF 120  
| | | | |  
QY 121 MVEDEETLHNPVYMGDGVLDODGTFIEELIKNVGKYGHDREGGFINDBI FVELVNLGQ 180  
| | | | |  
DB 121 MVEDEETLHNPVYMGDGVLDODGTFIEELIKNVGKYGHDREGGFINDBI FVELVNLGQ 180  
| | | | |  
QY 181 YNDDDDDDGDDPEREREKQDLDRHDHDKESRPPRPSPDKI FEAISNFPDQGTAEEL 240  
| | | | |  
DB 181 YNDDDDDDGDDPEREREKQDLDRHDHDKESRPPRPSPDKI FEAISNFPDQGTAEEL 240  
| | | | |  
QY 241 KEKTKELTEQQLPALPPECTPNIDGNVAVSQRSGLSHFTLFCRCRYDCFLAPFH 300  
| | | | |  
DB 241 KEKTKELTEQQLPALPPECTPNIDGNVAVSQRSGLSHFTLFCRCRYDCFLAPFH 300  
| | | | |  
QY 301 ATPVTYRKNTETALDNKPGSPQCYHLGAKGPAALLTABRTKTPPKRPGRRRGRLPN 360  
| | | | |  
DB 301 ATPVTYRKNTETALDNKPGSPQCYHLGAKGPAALLTABRTKTPPKRPGRRRGRLPN 360  
| | | | |  
QY 361 NSSRPSTPTINVLSEKQDSDREAGTGTGENDKBEKKDETSSSSKANSRCQPIKM 420  
| | | | |  
DB 361 NSSRPSTPTINVLSEKQDSDREAGTGTGENDKBEKKDETSSSSKANSRCQPIKM 420  
| | | | |  
QY 421 KPNTEPPENVWSGAESMFRVLIGTYDNFCALRLIGTTCQVYEFYKESSTIAPA 480  
| | | | |  
DB 421 KPNTEPPENVWSGAESMFRVLIGTYDNFCALRLIGTTCQVYEFYKESSTIAPA 480  
| | | | |  
QY 481 PABVDVPPRRKKRKHLLMAHGRKIQLKDGSSNHYVYTOPCDHPROPDSCSPCYIAQ 540  
| | | | |  
DB 481 PABVDVPPRRKKRKHLLMAHGRKIQLKDGSSNHYVYTOPCDHPROPDSCSPCYIAQ 540  
| | | | |  
QY 541 NFCEKFCQCSSECONRPFPGCRCAQCNCTKQPCYLAVRBCDPDLCTGADHMDSKNVS 600  
| | | | |  
DB 541 NFCEKFCQCSSECONRPFPGCRCAQCNCTKQPCYLAVRBCDPDLCTGADHMDSKNVS 600  
| | | | |  
QY 601 CKNCISIORGSKHLLAPSDVAGMGIFIKDPVOKNFISEYCGEIIISODEADRGKYDK 660  
| | | | |  
DB 601 CKNCISIORGSKHLLAPSDVAGMGIFIKDPVOKNFISEYCGEIIISODEADRGKYDK 660  
| | | | |  
QY 661 YMGCSFLPNLNDPVVDATRKGNKIRFANHSVNPCTAKV 699  
| | | | |  
DB 661 YMGCSFLPNLNDPVVDATRKGNKIRFANHSVNPCTAKV 699  
| | | | |  
RESULT 5  
Q571LS\_MOUSE PRELIMINARY; PRT; 779 AA.  
ID Q571LS\_MOUSE  
AC Q571LS\_MOUSE  
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
DT 10-MAY-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE MKIAA065 protein (fragment).  
GN Name=Bzh2; Synonyms=mkIAA065;

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP Tissue-embryonic tail;  
 RC Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hirakawa S.,  
 RA Suga Y., Nagase T., Ohara O., Koga H.;  
 RT Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene.  
 RT The Complete Nucleotide Sequences of Mouse KIAA-homologue cDNAs  
 RT Identified by Screening of Terminal sequences of cDNA Clones Randomly  
 RT Sampled from Size-Fractionated Libraries. "  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 DR EMBL; AK220174; BAD90359.1; -; mRNA.  
 DR Ensembl; ENSMUSG0000023687; Mus musculus.  
 DR MGI; MGI:107940; E2n2.  
 DR GO; GO:0045120; C:promote; IDA.  
 DR GO; GO:0003682; P:chromatin binding; IDA.  
 DR GO; GO:0042054; P:histone methyltransferase activity; IDA.  
 DR GO; GO:0005515; P:protein binding; IPI.  
 DR GO; GO:0016571; P:histone methylation; IDA.  
 DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.  
 DR InterPro; IPR001005; Myb\_DNA\_bd.  
 DR InterPro; IPR001214; SET.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00717; SANT; 2.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS50280; SET; 1.  
 DR NON TER 1  
 SQ SEQUENCE 779 AA; 88523 MW; A0CC1A16CA86E702 CRC64;  
 Query Match 98.6%; Score 3756; DB 2; Length 779;  
 Best Local Similarity 98.3%; Pred. No. 1.5e-211;  
 Matches 687; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MGQTGKSEKGVCKRKRKVKSEYMLRQLKFRADDEVSMSPSRKRLERTETLIDGM 60  
 DB 34 MGQTGKSEKGVCKRKRKVKSEYMLRQLKFRADDEVSMSPSRKRLERTETLIDGM 93  
 QY 61 KQRIQPVHILTVSSLRGTRECVSYSDLPFTQVYIPLKTLNANASVPIMWSPLQGNF 120  
 DB 94 KQRIQPVHILTVSSLRGTRECVSYSDLPFTQVYIPLKTLNANASVPIMWSPLQGNF 153  
 QY 121 MVEDETVLANIPYMGDEVLDQGTITBELIKKYDQKVGHDRCGFTINDRIFVELYNALGQ 180  
 DB 154 MVEDETVLANIPYMGDEVLDQGTITBELIKKYDQKVGHDRCGFTINDRIFVELYNALGQ 213  
 QY 181 YNDDDDDDGDDPDERBEKQKLEHRDCKESPRPKPSDIFPAISMPFDKGTAEEL 240  
 DB 214 YNDDDDDDGDDPDERBEKQKLEHRDCKESPRPKPSDIFPAISMPFDKGTAEEL 273  
 QY 241 KKKYKELTEQQLPGALPECTPNIDGPNAKSVQREQSLSFTLLRCRCFKYDCFLAPFH 300  
 DB 274 KKKYKELTEQQLPGALPECTPNIDGPNAKSVQREQSLSFTLLRCRCFKYDCFLAPFH 333  
 QY 301 ATPNTYKKNKTTALDNKPCGPOCYQH.BGAKFPAALTAERIKTPPKGRRGRRLPN 360  
 DB 334 ATPNTYKKNKTTALDNKPCGPOCYQH.BGAKFPAALTAERIKTPPKGRRGRRLPN 393  
 QY 361 NSSRSTPTINVLASGKQDSDREAGTENGGENNDEBEKKOETSSSSSANSRCQPTKM 420  
 DB 394 NSSRSTPTINVLASGKQDSDREAGTENGGENNDEBEKKOETSSSSSANSRCQPTKM 453  
 QY 421 KENIBPENWVGASGAFMFRVLIGTYDNCFAIRLIGTKICROVTERVRESIIAPA 480  
 DB 454 KENIBPENWVGASGAFMFRVLIGTYDNCFAIRLIGTKICROVTERVRESIIAPA 513

QY 481 PAEDVTPPRKKRGRRLMAAHCRIQLKKDSSNHNYPDCDHPROPCDSSCPCLVIAQ 540  
 DB 514 PREDVTPPRKKRGRRLMAAHCRIQLKKDSSNHNYPDCDHPROPCDSSCPCLVIAQ 573  
 QY 541 NPEKFCQCCSECCNRPFCRCCKAQCNTRKQCECYLAVERCDPDLCLTGAADHMSKNVS 600  
 DB 574 NPEKFCQCCSECCNRPFCRCCKAQCNTRKQCECYLAVERCDPDLCLTGAADHMSKNVS 633  
 QY 601 CKNCSIQGSKKHLILASDVAGNIFPKDPQKNEFISYCGEIIISQDEADRRKRYDK 660  
 DB 634 CKNCSIQGSKKHLILASDVAGNIFPKDPQKNEFISYCGEIIISQDEADRRKRYDK 693  
 QY 661 YMCSPFLNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
 DB 694 YMCSPFLNANDPVVDATRKGNKIRFANHSVNPNCYAKV 732  
 RESULT 6  
 E2H2\_MOUSE STANDARD; PRT; 746 AA.  
 AC Q61188; Q6R090; Integrated into UniProtKB/Swiss-Prot.  
 DT 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1996, sequence version 1.  
 DT 07-FEB-2006, entry version 48.  
 DE Enhancer of zeste homolog 2 (ENX-1).  
 GN Name=Ezh2; Synonyms=Bxh1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP Tissue-Placenta;  
 RC MEDLINE=97014262; PubMed=8861097; DOI=10.1016/0925-4773(96)00499-6;  
 RX Robert O., Sures I., Closek T., Fuchs M., Ullrich A.;  
 RA "Isolation and developmental expression analysis of Bxh-1, a novel  
 RT mouse Polycomb group gene.";  
 RL Mech. Dev. 55:171-184(1996).  
 [2]  
 RN NUCLEOTIDE SEQUENCE OF 134-497.  
 RP STRAIN=129/Sv.  
 RC MEDLINE=99160476; PubMed=10051331; DOI=10.1007/s003359900993;  
 RX Labile G., Haynes A.R., Lebersorger A., O'Carroll D., Mattei M.-G.,  
 RA Denny P., Brown S.D.M., Jenuwein T.;  
 RT "The murine polycomb-group genes ezh1 and ezh2 map close to hox gene  
 RT clusters on mouse chromosomes 11 and 6.";  
 RL Mamm. Genome 10:311-314(1999).  
 CC - FUNCTION: May be involved in the regulation of gene transcription  
 CC and chromatin structure.  
 CC - SUBUNIT: Binds ATRX via the SET domain (By similarity).  
 CC - SUBCELLULAR LOCATION: Nucleus (Probable).  
 CC - ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=ENX-1A;  
 CC IsoId=Q61188-1; Sequence=Displayed;  
 CC Name=ENX-1B;  
 CC IsoId=Q61188-2; Sequence=VSP\_001501;  
 CC - TISSUE SPECIFICITY: Widely expressed in early embryos. In later  
 CC embryogenesis, expression restricted to central and peripheral  
 CC nervous system, liver and thymus. In adult, highest expression in  
 CC spleen, testis and placenta. Lower levels in intestine and muscle  
 CC and very low levels in brain and liver. No expression in heart,  
 CC thyroid gland, lung and kidney.  
 CC - DEVELOPMENTAL STAGE: Expressed in both adult and embryo with  
 CC highest levels in early embryogenesis.  
 CC - SIMILARITY: Belongs to the E2 family.  
 CC - SIMILARITY: Contains 1 SET domain.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 DR EMBL; U52951; AAC52655.1; -; mRNA.

EMBL; AF104359; AAD54020.1; -, Genomic DNA.  
 DR HSPB; Q8X225; 1ML9.  
 DR Ensembl; ENSMUSG0000029687; Mus musculus.  
 DR MGI; MGI:107940; Ezh2.  
 DR LinkMap; 061188; -.  
 DR GO; GO:0045120; C:promoteus; IDA.  
 DR GO; GO:0003682; P:chromatin binding; IDA.  
 DR GO; GO:0042054; P:histone methyltransferase activity; IDA.  
 DR GO; GO:0005515; P:protein binding; IPI.  
 DR GO; GO:0016571; P:histone methylation; IDA.  
 DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.  
 DR InterPro; IPR001005; Myb\_DNA\_Bd.  
 DR InterPro; IPR001214; SET.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00717; SANT; 2.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS50280; SET; 1.  
 DR Alternative splicing; DNA-binding; Nuclear protein; Phosphorylation;  
 KW Transcription; Transcription regulation.  
 FT CHAIN 1 746  
 /FTid=PRO\_0000213993.  
 FT DOMAIN 611 731  
 /SET.  
 FT MOTIF 490 495  
 /Nuclear localization signal (Potential).  
 FT COMPBIA5 523 605  
 /Cys-rich.  
 FT MOD\_RBS 487 487  
 /Phosphothreonine (By similarity).  
 FT VARSPIC 511 553  
 /DSSNMVNYQPCDHPRQPCDSCPCVIAQNFCKFCQCSS  
 /EC -> G (in isoform ENK-1B).  
 /FTid=VSP\_001501.  
 FT CONFLICT 159 161  
 /Missing (In Ref. 2).  
 FT SEQUENCE 746 AA; 85336 MW; 0435C021963ED24E CRC64;  
 Query Match 159 161  
 Beat Local Similarity 98.1%; Score 3750; DB 1; Length 746;  
 Matches 686; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

DB 541 NFECKFCQSSBECNPFGRCKAQCNTKCCCTAABECDBDCLTCGAADHMDSKNVS 600  
 QY 601 CKNCISIGRSGKHLALAPSDVAGMGIFIKDPVQKNFISYCEGIIISQDEADRGKYDK 660  
 DB 601 CKNCISIGRSGKHLALAPSDVAGMGIFIKDPVQKNFISYCEGIIISQDEADRGKYDK 660  
 QY 661 YMCSPFLNANDPVYDATRGNKIRFANHSVNPCTYAKV 699  
 DB 661 YMCSPFLNANDPVYDATRGNKIRFANHSVNPCTYAKV 699  
 RESULT 7  
 Q372H6\_MOUSE PRELIMINARY; PRT; 746 AA.  
 AC Q372H6\_MOUSE  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 11-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Erythroblast cDNA, RIKEN full-length enriched library,  
 DE clone:K0C0026809 product:enhancer of zeste homolog 2 (Drosophila),  
 DE full insert sequence.  
 OS Name=Ezh2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RC STRAIN=C57BL/6J;  
 RC PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.B., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Winkler L.G., Altshuler S., Allen J.E.,  
 RA Amesel-Imphombato A., Apweiler R., Attwells R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christofidele A., Clutebuck D.R.,  
 RA Crowe M.L., Dalla B., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Furuki S., Gariboldi M.,  
 RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.B.,  
 RA Guetlich S., Harbers M., Hayashizaki Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummelbeck L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jait M., Kanapin A., Katoh M., Kawagawa Y., Kato J., Kitamura H.,  
 RA Kuroki H., Kollas G., Krishnan S.P., Kruger A., Kumerfeld S.K.,  
 RA Kurochkin I.V., Larreau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Faber S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Roest B., Ruan Y., Salzberg S.L., Sanderlin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Sempke C.A., Seno S., Seesla L., Sheng Y.,  
 RA Shiba Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugiyama K., Sultana R., Takenaka Y., Takai K.,  
 RA Tamojima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yang K.,  
 RA Yamanishi H., Zdobych E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Walthers C., Wiatlick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fujita S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imanura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Kono H., Nakano K., Nimmiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,



RA Tagami M., Waki K., Wataniki A., Okamura-Ohno Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC PubMed=16141073; DOI=10.1126/science.1112009;  
RG Riken Genome Exploration Research Group, and Genome Science Group  
RT (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense transcription in the mammalian transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Oseato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Brad D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Della E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmel S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kani A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kongsava A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numa K., Okido T., Pavan W.J., Pereira G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Raveli T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelin A., Schneider C., Sempé C.A., Seton M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilting L.G., Wymshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavoian M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hironaka-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
RA Yasunishi A., Yoshino Y., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,  
RA Kuehl P., Lewis S., Mateno Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stabill F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,  
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Saeki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 403:683-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,  
RA Kono H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanabe M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multichipillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RA Arikawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Kono H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Wataniki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC  
CC EMBL, AK157853; BAB34232.1; -; mRNA.  
DR GMI; MGI:107940; E2h2.  
DR GO; GO:0045120; C:promoter binding; IDA.  
DR GO; GO:0003682; F:chromatin binding; IDA.  
DR GO; GO:0042054; F:histone methyltransferase activity; IDA.  
DR GO; GO:0005515; F:protein binding; IPT.  
DR GO; GO:0016571; P:histone methylation; IDA.  
DR GO; GO:0051154; P:negative regulation of striated muscle cell. .; IDA.  
DR InterPro: IPR001005; MP DNA\_Bd.  
DR InterPro: IPR001214; SET.  
DR Pfam: PF00856; SET; 1.  
DR SMART: SM00717; SANT; 2.  
DR SMART: SM00317; SET; 1.  
DR PROSITE: PS0280; SET; 1.  
SQ SEQUENCE 746 AA; 85332 MW; BAD303CD9CB6FA4 CRC64;  
Query Match 98.4%; Score 3747; DB 2; Length 746;  
Best Local Similarity 98.1%; Pred. No. 4.9e-211;  
Matches 686; Conservative 7; Mismatches 6; Indels 0; Gaps 0;  
QY 1 MGQGGKSEKPKVCKRRKVKSEYVRLRLQKPRRADYKVSFSSNRQKILERTETILNDEM 60  
DB 1 MGQGGKSEKPKVCKRRKVKSEYVRLRLQKPRRADYKVSFSSNRQKILERTETILNDEM 60  
QY 61 KQRIQPVHILTSVSLGTRSCVTSDDLDEPPYVILPLKTLNAVASVIMYSWSPLOQNF 120  
DB 61 KQRIQPVHILTSVSLGTRSCVTSDDLDEPPYVILPLKTLNAVASVIMYSWSPLOQNF 120  
QY 121 MVEDETVLHNPYMGDEVLDDGTFPIELINQYDGKVGHGDECGGINDIEIVELVNALGQ 180  
DB 121 MVEDETVLHNPYMGDEVLDDGTFPIELINQYDGKVGHGDECGGINDIEIVELVNALGQ 180  
QY 181 YNDDDDDDGDDPBERREKQDLBPHRDCKSRPPRKPSPKIFPAISMPFDDGTABEL 240  
DB 181 YNDDDDDDGDDPBERREKQDLBPHRDCKSRPPRKPSPKIFPAISMPFDDGTABEL 240  
QY 241 KEKYEELTEQQLPGLPEPCTPNIDGPAKSVQREQSILSHPTLFCRCFCYDCLADPH 300  
DB 241 KEKYEELTEQQLPGLPEPCTPNIDGPAKSVQREQSILSHPTLFCRCFCYDCLADPH 300

QY 301 ATPTTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRLPN 360  
DB 301 ATPTTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRLPN 360  
QY 361 NSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCOTPIKM 420  
DB 361 NSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCOTPIKM 420  
QY 421 KPNLEPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESSIIAPA 480  
DB 421 KPNLEPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESSIIAPA 480  
QY 481 PAEDVDTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCPCTVIAQ 540  
DB 481 PTEVDVTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCPCTVIAQ 540  
QY 541 NFCEKFCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHDSKNVS 600  
DB 541 NFCEKFCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHDSKNVS 600  
QY 601 CKNGSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRGKXYDK 660  
DB 601 CKNGSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRGKXYDK 660  
QY 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
DB 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
RESULT 8  
Q6AXH7 MOUSE PRELIMINARY; PRT; 742 AA.  
ID 06AXH7-  
AC 06AXH7-  
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2004, sequence version 1.  
DT 07-FEB-2006, entry version 15.  
DB B2h2 protein.  
GN Name=B2h2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.V., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.R.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko I., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Goughwood J., Schmutz J., Myers R.M.,  
RA Butcherfield V.S.N., Krzywicki M.I., Skalska U., Smallus D.R.,  
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Director MGC Project;  
RL Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.  
CC

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC EMBL: BC079538; AAH79538.1; -, mRNA.  
DR Ensembl: ENSMUSG0000029687; Mus musculus.  
DR MGI: MGI:107940; B2h2.  
DR GO: GO:0045120; C:proncleus; IDA.  
DR GO: GO:0003682; F:chromatin binding; IDA.  
DR GO: GO:0042054; F:histone methyltransferase activity; IDA.  
DR GO: GO:0005515; F:protein binding; IPI.  
DR GO: GO:0016571; F:histone methylation; IDA.  
DR GO: GO:0051154; P:negative regulation of striated muscle cell. .; IDA.  
DR InterPro: IPR001005; Myb\_DNA\_Bd.  
DR InterPro: IPR001214; SET.  
DR Pfam: PF00856; SET; 1.  
DR SMART: SM00717; SANT; 2.  
DR SMART: SM00317; SET; 1.  
DR PROSITE: PS50280; SET; 1.  
SQ SEQUENCE 742 AA; 84934 MW; A9CDB80ACB2C6DEA CRC64;  
Query Match 96.6%; Score 3679; DB 2; Length 742;  
Best Local Similarity 96.2%; Pred. No. 4,7e-207;  
Matches 677; Conservative 7; Mismatches 6; Indels 14; Gaps 2;  
QY 1 MGOTGKSKGKPVWCRKRVSEYRRLQLRFRPADSVKMFSSNRKILERTLQEW 60  
DB 1 MGOTGKSKGKPVWCRKRVSEYRRLQLRFRPADSVKMFSSNRKILERTLQEW 60  
QY 61 KORIOPIVHLITSVSLRGTRSCVTSDDLFPNOVIELKTLNVAAPVIMYSPLQONF 120  
DB 61 KORIOPIVHLITSVSLRGTRSCVTSDDLFPNOVIELKTLNVAAPVIMYSPLQONF 120  
QY 121 MVEDETVLHNIPLYMGDEVLDQDTGTFIEELIKNTDGYKVGREGCFINDEIFVELVNLAQ 180  
DB 121 MVEDETVLHNIPLYMGDEVLDQDTGTFIEELIKNTDGYKVGREGCFINDEIFVELVNLAQ 180  
QY 112 MVEDETVLHNIPLYMGDEVLDQDTGTFIEELIKNTDGYKVGREGCFINDEIFVELVNLAQ 171  
DB 112 MVEDETVLHNIPLYMGDEVLDQDTGTFIEELIKNTDGYKVGREGCFINDEIFVELVNLAQ 171  
QY 181 YNDDDDDDDDPERRERKQKLEDDHDDKESRPRRFPBDKIFEALISSMPDGTABEL 240  
DB 181 YNDDDDDDDDPERRERKQKLEDDHDDKESRPRRFPBDKIFEALISSMPDGTABEL 240  
QY 172 YNDDDDDDDDPERRERKQKLEDDHDDKESRPRRFPBDKIFEALISSMPDGTABEL 231  
DB 172 YNDDDDDDDDPERRERKQKLEDDHDDKESRPRRFPBDKIFEALISSMPDGTABEL 231  
QY 241 KEKTKELTEQQLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCRCFYDCFLH--- 297  
DB 241 KEKTKELTEQQLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCRCFYDCFLH--- 297  
QY 232 KEKTKELTEQQLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCRCFYDCFLH--- 291  
DB 232 KEKTKELTEQQLPGALPPECTPNIDGNPAKSVQREGSLHSFHTLFCRCRCFYDCFLH--- 291  
QY 298 --PHATPNTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRL 355  
DB 298 --PHATPNTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRL 355  
QY 292 SYSFHAATPNTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRL 351  
DB 292 SYSFHAATPNTYKAKNTETALDNKPCGPOCYOHLGAKBFAALTAERIKTPPKRGGRRL 351  
QY 356 GRLPNSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCQ 415  
DB 356 GRLPNSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCQ 415  
QY 352 GRLPNSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCQ 411  
DB 352 GRLPNSSRPSTPTINVLASKOTDSREAGTETGGENNDKEBEKKODETSSSSRANRCQ 411  
QY 416 TPIPKKPIRPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESS 475  
DB 416 TPIPKKPIRPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESS 475  
QY 412 TPIPKKPIRPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESS 471  
DB 412 TPIPKKPIRPPENVMGSAEMFVILGTYNDNFCALIALIGTKICQVYEFVYESS 471  
QY 476 IIAAPADVDVTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCP 535  
DB 476 IIAAPADVDVTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCP 535  
QY 472 IIAAPADVDVTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCP 531  
DB 472 IIAAPADVDVTPPRKKRKHRLMAAHCRIQLKKGSSNHYVNYQPCDHPROPCDSSCP 531  
QY 536 CVIANKPCEKCCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHMD 595  
DB 536 CVIANKPCEKCCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHMD 595  
QY 532 CVIANKPCEKCCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHMD 591  
DB 532 CVIANKPCEKCCOCCSSBQONRPFQCGCKAQCNTPKOCPTILANRCBPDLCTLTCGAADHMD 591  
QY 596 SKNYSCKKCSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRG 655  
DB 596 SKNYSCKKCSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRG 655  
QY 592 SKNYSCKKCSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRG 651  
DB 592 SKNYSCKKCSIQSGSKKHLIAPSDVAGMGIFIKDPVQKNFISEYCSGELISQDEADRG 651  
QY 656 KYVDKYMCSFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
DB 656 KYVDKYMCSFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
CC

RESULT 9  
OSX80\_XENLA PRELIMINARY; PRT; 748 AA.  
ID OSX80;  
AC OSX80;  
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.  
DT 23-NOV-2004, sequence version 1.  
DT 21-FEB-2006, entry version 15.  
DE B2h2 protein.  
GN Name=B2h2;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
NCBI\_TaxID=8355;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Kleim S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT Initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
[2]  
NP NUCLEOTIDE SEQUENCE.  
RP TISSUE-Ovary;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold R.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hejblum P.,  
RA Ditschenko L., Marzetta K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Mazy D.M., Sodergren B.U., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield V.S.N., Krzywinski M.T., Skalska U., Smalins D.B.,  
RA Scheraga A., Schein J.E., Jones S.J.W., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
NP NUCLEOTIDE SEQUENCE.  
RP TISSUE-Ovary;  
RC TISSUE-Ovary;  
RA Klein S., Gethard D.S.;  
RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC  
EMBL: BC084193; AA084193.1; -; mRNA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR01005; Myb\_DNA\_bd.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00717; SANT; 2.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PSS0280; SET; 1.  
SQ SEQUENCE 748 AA; 8555 MW; 2916A4B3CB8C4E CRC64;  
Query Match 93.6%; Score 3566; DB 2; Length 748;  
Best Local Similarity 93.0%; Pred. No. 2e-200;  
Matches 652; Conservative 25; Mismatches 22; Indels 2; Gaps 1;  
QY 1 MGQTGKSKGKGVCKRKRKVRKYSYMLRLQKRRADAEVKSMTNTRQKIMERTETLLNQEW 60  
|||||

DB 1 MGQTGKSKGKGVCKRKRKVRKYSYMLRLQKRRADAEVKSMTNTRQKIMERTETLLNQEW 60  
QY 61 KQRIOPVHITSSSLRGTRGCVSTDDPPVOVPLKTLNAAVASVIMVSWGLQONF 120  
DB 61 KQRIOPVHITSSSLRGTRGCVSTDDPPVOVPLKTLNAAVASVIMVSWGLQONF 120  
QY 121 MVEDETVLHNPVNGDEVLDDGGTFIEELINQYQKYNHDEBCGFINDIEFVELNALGQ 180  
DB 121 MVEDETVLHNPVNGDEVLDDGGTFIEELINQYQKYNHDEBCGFINDIEFVELNALGQ 180  
QY 181 YNDDDDDDGGD--PERRRKKQDLBDRDDKSRPPKPSDXTFEALISSMPDQGTAE 238  
DB 181 YSDYEDDEDGDDNDQDDEDDTAQDDNMDKEXQPLKFPSSDKTPEALISSMPDQGTSE 240  
QY 239 ELKRTKTELTEQQLPGALPECTNTINGPNKAKSYQREOSLSFRTLPCRRCFKTDCFLHP 298  
DB 241 ELKRTKTELTEQQLPGALPECTNTINGPNKAKSYQREOSLSFRTLPCRRCFKTDCFLHP 300  
QY 299 PHATPNYTKRKNTETALDNKPCGPQCYOHLGAKFPAAALTAERIKTPPKRPGRRRGRRL 358  
DB 301 PHATPNYTKRKNTETALDNKPCGPQCYOHLGAKFPAAALTAERIKTPPKRPGRRRGRRL 360  
QY 359 PNNSSRSTPTINVLBSKDTSDRBAETGEGENNDKREBKODETSSSSRANRCQTP 418  
DB 361 PNNSSRSTPTINVLBSKDTSDRBAETGEGENNDKREBKODETSSSSRANRCQTP 420  
QY 419 KMKNNIPEPVENVSGABSMFRLIGTYDNPCLARLITKTCROVYERYVSSIIA 478  
DB 421 KMKNNIPEPVENVSGABSMFRLIGTYDNPCLARLITKTCROVYERYVSSIIA 480  
QY 479 PAPADVDTPPRKKGRKRRMAHCRKIQLKQDSSNNVNYQCDHPROPDSCPCVI 538  
DB 481 PVLEADVDTPPRKKGRKRRMAHCRKIQLKQDSSNNVNYQCDHPROPDSCPCVI 540  
QY 539 AQNCEKRCQCCSSCONRFPQCRCAQONTQPCYLAVRECDPLCTCGAADHWDKSN 598  
DB 541 AQNCEKRCQCCSSCONRFPQCRCAQONTQPCYLAVRECDPLCTCGAADHWDKSN 600  
QY 599 VSCNKGSIQRGSKKHLILAPSDVAGMGIFIKDPVQKNFISEYGEIISQDEADRGKVV 658  
DB 601 VSCNKGSIQRGSKKHLILAPSDVAGMGIFINDVQKNFISEYGEIISQDEADRGKVV 660  
QY 659 DKWVCSEFLFNNDPVVDATRKGNKIRPANHNVNPNCAKY 699  
DB 661 DKWVCSEFLFNNDPVVDATRKGNKIRPANHNVNPNCAKY 701  
RESULT 10  
O98SM3\_XENLA PRELIMINARY; PRT; 748 AA.  
ID O98SM3;  
AC O98SM3;  
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 20.  
DE Enhancer of zeste.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
NCBI\_TaxID=8355;  
[1]  
NP NUCLEOTIDE SEQUENCE.  
RP MEDLINE=21184113; PubMed=11287189; DOI=10.1016/S0925-4773(01)00304-5;  
RA Barnett M.W., Seville R.A., Nijjar S., Old R.W., Jones B.A.;  
RT "Xenopus Enhancer of Zeste (XEZ), an anteriorly restricted polycomb  
RT gene with a role in neural patterning.";  
RL Mech. Dev. 102:157-167(2001).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC  
EMBL: AF351126; AA030208.1; -; mRNA.  
DR HSPD; Q8X225; IML9.





NCBI\_TaxID=9606;

OX [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;

RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,

RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,

RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,

RA Fowell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,

RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,

RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,

RA Ozeresky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,

RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,

RA Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,

RA Tin-Wollam A.-M., Abbott A., Mink P., Maupin R., Strommatt C.,

RA Latreille P., Miller N., Johnson D., Murray J., Moesner J.P.,

RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,

RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,

RA Hickenbotham M.T., Bleded J., Williams D., Bedell J.A., Mardis E.R.,

RA Clifton S.W., Chisoe S.L., Marra M.A., Raymond C., Haugen B.,

RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubbs K.,

RA Sims B., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,

RA Baertsch R.A., Brent M.R., Keibler E., Flicker P., Bork P., Suyama M.,

RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,

RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.B., Green E.D.,

RA Waterston R.H., Wilson R.K.;

RT "The DNA sequence of human chromosome 7.";

RL Nature 424:157-164(2003).

[2]

RP NUCLEOTIDE SEQUENCE.

RA Bauer C., Holmes A., Mead K.;

RT "The sequence of Homo sapiens PAC clone RP5-1151M5.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

[3]

RP NUCLEOTIDE SEQUENCE.

RA Waterston R.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

[4]

RP NUCLEOTIDE SEQUENCE.

RA Wilson R.;

RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonCommercial license

CC -----

CC EMBL: AC006323; AAS02035.1; -; Genomic DNA.

CC Ensembl: ENSG00000106462; Homo sapiens.

DR GO: GO:0005634; C:nucleus, IEA.

DR GO: GO:0005677; F:DNA binding, IEA.

DR InterPro: IPR001005; MYD\_DNA\_Bd.

DR InterPro: IPR001214; SET.

DR Pfam: PF00856; SET, 1.

DR SMART: SM00317; SANT, 2.

DR PROSITE: PS00280; SET, 1.

DR Hypothetical protein.

FT NON TER 1

SEQUENCE 664 AA; 75478 MW; 5E5FCB1B5D8651C8 CRC64;

Query Match 89.0%; Score 3388; DB 2; Length 664;

Best Local Similarity 100.0%; Fred. No. 4.6e-190; Indels 0; Gaps 0;

Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 CSVTSDLPPTVOVPLKLTINAVASVIMYSMSPLQONFEVDEYETLANIPYMGDEVLLDOD 142

DB 83 CSVTSDLPPTVOVPLKLTINAVASVIMYSMSPLQONFEVDEYETLANIPYMGDEVLLDOD 142

DB 1 CSVTSDLPPTVOVPLKLTINAVASVIMYSMSPLQONFEVDEYETLANIPYMGDEVLLDOD 60

DB 143 GTFIBELIKNYDGKVGAGRECGFINDBIFVELVNLGYQVNDDDDDDDGDDPEEREKQND 202

DB 61 GTFIBELIKNYDGKVGAGRECGFINDBIFVELVNLGYQVNDDDDDDDGDDPEEREKQND 120

DB 203 LEDHDDKSRPRPKRPPSKITFEALSSMPDGTAEELKXKXKELTEOOLPGALPECP 262

DB 121 LEDHDDKSRPRPKRPPSKITFEALSSMPDGTAEELKXKXKELTEOOLPGALPECP 180

QY 263 NIDGNPAKSVQREQSLSHFTLLFCRRCFKDCFLHPHATPTNTYKKNITETALDNKRCGP 322

DB 181 NIDGNPAKSVQREQSLSHFTLLFCRRCFKDCFLHPHATPTNTYKKNITETALDNKRCGP 240

QY 323 QCVQHGAEPFAALTAERIKTPPKRPGRRBGRPLPNSSSRPSTPTINYLESKDTSDR 382

DB 241 QCVQHGAEPFAALTAERIKTPPKRPGRRBGRPLPNSSSRPSTPTINYLESKDTSDR 300

QY 383 EAGTETGENDXEEKEKDETSSEANSRCQPTIKMKENIBPENVEWSGAEMFRV 442

DB 301 EAGTETGENDXEEKEKDETSSEANSRCQPTIKMKENIBPENVEWSGAEMFRV 360

QY 443 LIGTYNFCALARLITKTCROYERVRVSSIIAPAPAEVDTPPRKKRGRMAAH 502

DB 361 LIGTYNFCALARLITKTCROYERVRVSSIIAPAPAEVDTPPRKKRGRMAAH 420

QY 503 CRKQLKDDSSNNVNYOPCDHPROPDSSCPVIAQNFCEKFCQSSSECONRFPQCRC 562

DB 421 CRKQLKDDSSNNVNYOPCDHPROPDSSCPVIAQNFCEKFCQSSSECONRFPQCRC 480

QY 563 KAQNTKQCPYLAVERCDPDLCTCGAADHMSKNVSCNRCISIQSGSKHLLAPSDVA 622

DB 481 KAQNTKQCPYLAVERCDPDLCTCGAADHMSKNVSCNRCISIQSGSKHLLAPSDVA 540

QY 623 GWGIFIDPVQKNFEISYCGEIIISQDEADRGKVTYKMCSTFLPNLNDPVYDATKGN 682

DB 541 GWGIFIDPVQKNFEISYCGEIIISQDEADRGKVTYKMCSTFLPNLNDPVYDATKGN 600

QY 683 KIRFANHSVNPNCYAKV 699

DB 601 KIRFANHSVNPNCYAKV 617

RESULT 14

075M00 HUMAN PRELIMINARY; PRT; 669 AA.

AC 075M00;

DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 05-JUL-2004, sequence version 1.

DT 21-FEB-2006, entry version 10.

DE Hypothetical protein EZH2 (Fragment).

GN Name=EZH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;

OC Homo.

OX NCBI\_TaxID=9606;

FN [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;

RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,

RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,

RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,

RA Fowell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,

RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,

RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,

RA Ozeresky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,

RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,

RA Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,

RA Tin-Wollam A.-M., Abbott A., Mink P., Maupin R., Strommatt C.,

RA Latreille P., Miller N., Johnson D., Murray J., Moesner J.P.,

RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,

RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,

RA Hickenbotham M.T., Bleded J., Williams D., Bedell J.A., Mardis E.R.,

RA Clifton S.W., Chisoe S.L., Marra M.A., Raymond C., Haugen B.,

RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubbs K.,

RA Sims B., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,

RA Baertsch R.A., Brent M.R., Keibler E., Flicker P., Bork P., Suyama M.,

RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,

RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.B., Green E.D.,

RA Waterston R.H., Wilson R.K.;

RT "The DNA sequence of human chromosome 7.";

RL Nature 424:157-164 (2003).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Bauer C., Holmes A., Mead K.;  
 RT "The sequence of Homo sapiens PAC clone RP5-1151M5."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Waterston K.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Wilson R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC distributed under the Creative Commons Attribution-NonDerivs license  
 CC -----  
 DR EMBL: AC006323; AAC02036.1; -; Genomic\_DNA.  
 DR Ensembl: ENSG00000106462; Homo sapiens.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR InterPro: IPR001005; Myb\_DNA\_Bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS50280; SET; 1.  
 DR Hypothetical protein.  
 KM  
 FT NON TER 1  
 SQ SEQUENCE 669 AA; 76132 MW; 413165183ABBF7D1 CRC64;  
 Query Match 88.4%; Score 3367.5; DB 2; Length 669;  
 Best Local Similarity 99.0%; Pred. No. 7,3e-189;  
 Matches 616; Conservative 0; Mismatches 1; Indels 5; Gaps 1;  
 QY 83 CSVSDLPPTQVITLKLINAVSPIMYSPLQONFVBEVLAHNPYMGDEVLD 142  
 DB 1 CSVSDLPPTQVITLKLINAVSPIMYSPLQONFVBEVLAHNPYMGDEVLD 60  
 QY 143 GTFIEELIKNYDKVHGRCGFINDEIFVELVNALGQVNDDDDDDDGDPREBEKOD 202  
 DB 61 GTFIEELIKNYDKVHGRCGFINDEIFVELVNALGQVNDDDDDDDGDPREBEKOD 120  
 QY 203 LEDHDDKESRPRPRPSDKIFPAISMPDGTABELKXKYLTEQQLPGALPPECTP 262  
 DB 121 LEDHDDKESRPRPRPSDKIFPAISMPDGTABELKXKYLTEQQLPGALPPECTP 180  
 QY 263 NIDGNASVQREGLSHFHTLFCRCRKYDCLF-----PHATPNYKKNTEALDN 317  
 DB 181 NIDGNASVQREGLSHFHTLFCRCRKYDCLF-----PHATPNYKKNTEALDN 240  
 QY 318 KPCGQCTOHLGAGEPAAALTAERIKTPPKRPGRRGRPLNNSSRSTPTINVLASKD 377  
 DB 241 KPCGQCTOHLGAGEPAAALTAERIKTPPKRPGRRGRPLNNSSRSTPTINVLASKD 300  
 QY 378 TDSDEAGCTEGENNDKBEBEKODTSSSSSEANSRCOTPIKKNIPPEVENSAGA 437  
 DB 301 TDSDEAGCTEGENNDKBEBEKODTSSSSSEANSRCOTPIKKNIPPEVENSAGA 360  
 QY 438 SMFRLIGTYDNPALRIIGTKCRQYERFRVSSIIAPABEDVDTPRKKRGR 497  
 DB 361 SMFRLIGTYDNPALRIIGTKCRQYERFRVSSIIAPABEDVDTPRKKRGR 420  
 QY 498 LMAACRKLQKQSSNNHVNYPGCDHPROPCCSSCPVLAQNCEKFCQSSCCQNR 557  
 DB 421 LMAACRKLQKQSSNNHVNYPGCDHPROPCCSSCPVLAQNCEKFCQSSCCQNR 480  
 QY 558 PGCRCACQNTKQPCYLAVERCDPLCTGGAADHMSKNVSCNCSIQRSGKHLILA 617  
 DB 481 PGCRCACQNTKQPCYLAVERCDPLCTGGAADHMSKNVSCNCSIQRSGKHLILA 540  
 QY 618 PSDVAGKGIPIKDPQKKEPISBYCGRIISQDEADRRGKVDKMCSTLFNLNDFVDA 677

DB 541 PSDVAGKGIPIKDPQKKEPISBYCGRIISQDEADRRGKVDKMCSTLFNLNDFVDA 600  
 QY 678 TRKGNKIRFANHSVNPNCYATV 699  
 DB 601 TRKGNKIRFANHSVNPNCYATV 622  
 RESULT 15  
 Q90WP4\_TESTNG PRELIMINARY; PRT; 759 AA.  
 AC Q90WP4;  
 DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-DEC-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 20.  
 DE E2H2 Homolog.  
 GN Name=e2h2;  
 OS Tetradon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetradon.  
 OC NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Blood;  
 RX MEDLINE=22269943; Pubmed=12368471; DOI=10.1073/pnas.202284199;  
 RA Dasliya C., Hadji H., Ozouf-Costaz C., Nicoud S., Jallion O.,  
 RA Weissenbach J., Roest Croillat H.;  
 RT "Remarkable compartmentalization of transposable elements and  
 RT pseudogenes in the heterochromatin of the Tetradon nigroviridis  
 RT genome."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13636-13641 (2002).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC distributed under the Creative Commons Attribution-NonDerivs license  
 CC -----  
 DR EMBL: AJ313481; CAC86146.1; -; Genomic\_DNA.  
 DR HSSP; Q8X225; 1MD9.  
 DR Ensembl: GSTENG0007024001; Tetradon nigroviridis.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR InterPro: IPR001005; Myb\_DNA\_Bd.  
 DR InterPro: IPR001214; SET.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00717; SANT; 2.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS50280; SET; 1.  
 DR SEQUENCE 759 AA; 86834 MW; DBA5DFE76922F640 CRC64;  
 Query Match 83.1%; Score 3164.5; DB 2; Length 759;  
 Best Local Similarity 82.2%; Pred. No. 6,5e-177;  
 Matches 587; Conservative 53; Mismatches 51; Indels 23; Gaps 10;  
 QY 4 TGKSKGPGVWRKRVSEVRLQKRPRAADVKSNNRQKILERTILNOMKOR 63  
 DB 4 TGKSKGPGVWRKRVSEVRLQKRPRAADVKSNNRQKILERTILNOMKOR 63  
 QY 64 RIQPHILTSVSRGTRECSVTSDL--DEPTQVITLKLINAVSPIMYSPLQONFV 122  
 DB 64 RIQPHILTSVSRGTRECSVTSDL--DEPTQVITLKLINAVSPIMYSPLQONFV 123  
 QY 123 EDETVLHNPYMGDEVLDQGTPIBELIKNYDKVHGRCGFINDEIFVELVNALGQYN 182  
 DB 124 EDETVLHNPYMGDEVLDQGTPIBELIKNYDKVHGRCGFINDEIFVELVNALGQYS 183  
 QY 183 DDDDDDDGDDDEBEBEK-----QKLEDDHDD-----KESRP-----PKRPSDKIFPAIS 228  
 DB 184 DNEDEBEBEKDDFVYDKIDLCDSSKHPDGRKDLVNNBGSADGSKKPPSDKIFPAIS 243  
 QY 229 SMPDGTABELKXKYLTEQQLPGALPPECTPNIDGNASVQREGLSHFHTLFCRR 288  
 DB 244 AMPDGTABELKXKYLTEQQLPGALPPECTPNIDGNASVQREGLSHFHTLFCRR 303



```

QY 289 CFKXDCFLPHPHATPNTYKRAKNTETALDNKPCGPQYQHL--EG-AKEFAALTAERIXT 345
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 CFKXDCFLPHPHATPNTYKRAKNTETALDNKPCGPQYQHL--EG-AKEFAALTAERIXT 361
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 346 PPKRPGRRGRRLPNNSRRPSTPTINVLKSDREAGTETGGANNKKEERKQETS 405
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 PSKRPAGRRRGRQPNSSSRPSTPTVS--SESKADSDRE--GSKEDERENKDEDDKQENT 419
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 406 SSSBANSCOTPIWKPNIEEPENWESGAESMERVLIGTYDNPCAIARLIGTKCRQ 465
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 SSSBGNRCQTPVKKK-LVGAZBAVDWNGABASLFRVLIGTYDNPCAIARLIGTKCRQ 478
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 466 VYEFVVKESIIAPAPAEVDVTPPKKKRKHRLMAHCRKIQLKKDSSNHVINYQPCDH 525
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 VYEFVVKESIIAPAPAEVDVTPPKKKRKHRLMAHCRKIQLKKDSSNHVINYQPCDH 538
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 526 PRQPCDSSCPVIAQNFCEKFCQCSSECQNRPGCRCAQCNTKQCPCTLAVRBCDPDLQ 585
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 PRQPCDSSCPVIAQNFCEKFCQCSSECQNRPGCRCAQCNTKQCPCTLAVRBCDPDLQ 598
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 586 LTCGAADHWSKNSCKNCSIORGAKKHLILAPSDVAGMGIPIQDVOXNEFISYGEI 645
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 LTCGAADHWSKNSCKNCSIORGAKKHLILAPSDVAGMGIPIQDVOXNEFISYGEI 658
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 646 ISODEADRRGKVDKYMCSFLFNLNDFVVDATRKGNKIRPANHSVNPNCYAKV 699
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 ISODEADRRGKVDKYMCSFLFNLNDFVVDATRKGNKIRPANHSVNPNCYAKV 712
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: August 11, 2006, 22:28:04  
 Job time : 189.206 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:28:25 ; Search time 34.8291 Seconds  
(without alignments)  
1756.691 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGOTGKSKSGPVCKRKRKRVK.....KGNKIRFANHSVNPCTAKV 699

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0\*

Maximum Match 100\*

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/5\_COMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/6\_COMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/7\_COMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/H\_COMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/PCVTUS\_COMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/RE\_COMB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/Ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	746	2	US-09-589-892B-2
2	3569.5	93.7	707	2	US-10-104-047-3162
3	1934.5	50.8	760	2	US-09-589-892B-11
4	719.5	18.9	856	2	US-09-699-266A-13
5	679.5	17.8	902	2	US-09-699-266A-12
6	665.1	17.5	921	2	US-09-699-266A-9
7	612.1	16.1	689	2	US-09-177-249-2
8	612.1	16.1	689	2	US-09-061-769A-2
9	612.1	16.1	689	2	US-09-812-283-2
10	612.1	16.1	689	2	US-10-176-884-14
11	612.1	16.1	689	3	US-09-071-838A-2
12	612.1	16.1	932	2	US-10-176-884-15
13	470.5	12.4	359	2	US-09-699-266A-11
14	293	7.7	114	2	US-10-200-012-39
15	251	6.6	115	2	US-10-176-884-19
16	228	6.0	3969	2	US-08-061-376-5
17	228	6.0	3969	2	US-09-538-092-1262
18	227	6.0	86	2	US-09-640-211A-1182
19	227	6.0	86	2	US-09-640-211A-1182
20	222	5.8	477	2	US-09-876-221-2
21	210	5.5	883	2	US-09-248-796A-18931
22	209	5.5	1058	2	US-10-094-749-1909
23	204	5.4	412	2	US-09-589-892B-4
24	201	5.3	1016	2	US-09-949-016-11018
25	200	5.3	429	2	US-09-589-892B-16
26	196	5.1	192	2	US-09-270-767-32394

27	191.5	5.0	542	2	US-09-248-796A-15128	Sequence 15128, A
28	181	4.8	133	2	US-09-589-892B-14	Sequence 14, Appl
29	178.5	4.7	210	1	US-08-320-559-4	Sequence 4, Appl1
30	178.5	4.7	210	1	US-08-327-392-4	Sequence 4, Appl1
31	178.5	4.7	210	2	US-08-545-860D-4	Sequence 4, Appl1
32	178.5	4.7	210	5	PCT-US94-04496-4	Sequence 4, Appl1
33	177.5	4.7	134	2	US-09-589-892B-12	Sequence 12, Appl
34	174.5	4.6	117	2	US-10-200-012-30	Sequence 30, Appl
35	169.5	4.5	132	2	US-09-589-892B-13	Sequence 13, Appl
36	168.5	4.4	4019	2	US-09-854-133-425	Sequence 425, App
37	165	4.3	136	2	US-09-589-892B-15	Sequence 15, Appl
38	164	4.3	202	2	US-09-270-767-31876	Sequence 31876, A
39	164	4.3	202	2	US-09-270-767-47093	Sequence 47093, A
40	163	4.3	119	2	US-10-200-012-31	Sequence 31, Appl
41	160.5	4.2	62	2	US-09-177-249-82	Sequence 82, Appl
42	160.5	4.2	62	2	US-09-812-283-82	Sequence 82, Appl
43	160.5	4.2	62	3	US-09-071-838A-82	Sequence 82, Appl
44	149	3.9	89	2	US-10-200-012-32	Sequence 32, Appl
45	145	3.8	93	2	US-09-640-211A-975	Sequence 975, App

## ALIGNMENTS

RESULT 1						
US-09-589-892B-2						
Sequence 2, Application US/09589892B						
Patent No. 6689583						
GENERAL INFORMATION:						
APPLICANT: Jenuwein, Thomas						
APPLICANT: Laible, Gotz						
APPLICANT: O'Carroll, Donal						
APPLICANT: Bismaher, Frank						
APPLICANT: Rea, Stephen						
TITLE OF INVENTION: Chromatin-Regulator Genes						
FILE REFERENCE: 0652.1670001						
CURRENT APPLICATION NUMBER: US/09/589, 892B						
CURRENT FILING DATE: 2000-06-09						
PRIOR APPLICATION NUMBER: US 08/945, 988						
PRIOR FILING DATE: 1997-11-10						
PRIOR APPLICATION NUMBER: PCT/EP96/01818						
PRIOR FILING DATE: 1996-05-02						
PRIOR APPLICATION NUMBER: DE 195 16 776.7						
PRIOR FILING DATE: 1995-05-10						
NUMBER OF SEQ ID NOS: 21						
SOFTWARE: PatentIn version 3.2						
SEQ ID NO 2						
LENGTH: 746						
TYPE: PRT						
ORGANISM: Homo sapiens						
US-09-589-892B-2						
Query Match						
Best Local Similarity 100.0%, Pred. No. 0;						
Matches 699; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	MGOTGKSKSGPVCKRKRKRVK	100.0%	Score 3808;	DB 2;	Length 746;
DB	1	MGOTGKSKSGPVCKRKRKRVK				
QY	61	KORRQPHILITVSLSAGTRECSTSLDPPQYVITPLKTNAAVSVPMYWSLQONF				
DB	61	KORRQPHILITVSLSAGTRECSTSLDPPQYVITPLKTNAAVSVPMYWSLQONF				
QY	121	MVEDETVLANTPYMGDEVLDODGTFIEELIKNYDGKVGADRECGFINDEIFVELYNALGQ				
DB	121	MVEDETVLANTPYMGDEVLDODGTFIEELIKNYDGKVGADRECGFINDEIFVELYNALGQ				
QY	181	YNDDDDDDDGDDPEERBEKQKDLBDHDKESRPPKPPSDKIFPAISSMPPDKGTAEEL				
DB	181	YNDDDDDDDGDDPEERBEKQKDLBDHDKESRPPKPPSDKIFPAISSMPPDKGTAEEL				
QY	241	KERYELTEQQLPGLAPPECTPNIDGPAKSVQBRQSLASHFTTLFCRRCFPKDCIAPHFH				
DB	241	KERYELTEQQLPGLAPPECTPNIDGPAKSVQBRQSLASHFTTLFCRRCFPKDCIAPHFH				

Db 241 KKKYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLFCRCFCYDCFLHFFH 300  
Qy 301 ATPTTYRKQKTEETALDNKPCGPOCYOHLGAKERPAALTAARIKTPPRPGRRGRBLPN 360  
Db 301 ATPTTYRKQKTEETALDNKPCGPOCYOHLGAKERPAALTAARIKTPPRPGRRGRBLPN 360  
Qy 361 NSSRPTPTINVLBSKQDTSDBRAGTGTGENDKEEKKDETSSSSEANSRCQTP1KM 420  
Db 361 NSSRPTPTINVLBSKQDTSDBRAGTGTGENDKEEKKDETSSSSEANSRCQTP1KM 420  
Qy 421 KPNIEPPENWMSGASMSFRVLIGTYDNFCAIARLIGTKTCQVYEFYKSSIIAPA 480  
Db 421 KPNIEPPENWMSGASMSFRVLIGTYDNFCAIARLIGTKTCQVYEFYKSSIIAPA 480  
Qy 481 PABVDTPPRKKRKHRLMAHCKRIQLKDGSSNHVYNYQPCDHPRQPCDSCPCVIAQ 540  
Db 481 PABVDTPPRKKRKHRLMAHCKRIQLKDGSSNHVYNYQPCDHPRQPCDSCPCVIAQ 540  
Qy 541 NFCEKFCQCSSECCNRPFGCRCKAQCNTKQCPCYLAVERCDPDLCTGGAADHWSKNVS 600  
Db 541 NFCEKFCQCSSECCNRPFGCRCKAQCNTKQCPCYLAVERCDPDLCTGGAADHWSKNVS 600  
Qy 601 CKNCISIGSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISODEADRGRKYDK 660  
Db 601 CKNCISIGSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISODEADRGRKYDK 660  
Qy 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
Db 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699

## RESULT 2

US-10-104-047-3162  
Sequence 3162, Application US/10104047  
Patent No. 6943241  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. 6943241el full1 length cdna  
FILE REFERENCE: HI-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
PRIOR FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3162  
LENGTH: 707  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-3162

Query Match 93.7%; Score 3569.5; DB 2; Length 707;  
Best Local Similarity 94.3%; Pred. No. 1.9e-301;  
Matches 659; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

Qy 1 MGQTKSEKSEKGVCKRKKVSEYMLRQLKFRPADYKVSFSSSRQKILERTTEILNDQM 60  
Db 1 MGQTKSEKSEKGVCKRKKVSEYMLRQLKFRPADYKVSFSSSRQKILERTTEILNDQM 60  
Qy 61 KQRIQPVHILTSVSLRGTRCSVTSLDLPTQVILPKTLNVAASVDMYSWSPLQONF 120  
Db 61 KQRIQPVHILTSVSLRGTRCSVTSLDLPTQVILPKTLNVAASVDMYSWSPLQONF 120  
Qy 121 MWBETVLANIPYMGDEVLDQDGTPIBELIKNYDGKVGDRPCGFINDEIFVELVNALGQ 180  
Db 83 -VEBETVLANIPYMGDEVLDQDGTPIBELIKNYDGKVGDRPCGFINDEIFVELVNALGQ 141  
Qy 181 YNDDDDDGDDPESEREKQKLEJHRDDKESRPKPTPSDKIFPAISMPFDKGTAEEL 240  
Db 142 YNDDDDDGDDPESEREKQKLEJHRDDKESRPKPTPSDKIFPAISMPFDKGTAEEL 201  
Qy 241 KKKYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLFCRCFCYDCFLHFFH 300

Db 202 KKKYKELTEQOLPGALPPECTPNIDGNPAKSVQREQSLSHFTLFCRCFCYDCFLHFFH 261  
Qy 301 ATPTTYRKQKTEETALDNKPCGPOCYOHLGAKERPAALTAARIKTPPRPGRRGRBLPN 360  
Db 262 ATPTTYRKQKTEETALDNKPCGPOCYOHLGAKERPAALTAARIKTPPRPGRRGRBLPN 321  
Qy 361 NSSRPTPTINVLBSKQDTSDBRAGTGTGENDKEEKKDETSSSSEANSRCQTP1KM 420  
Db 361 NSSRPTPTINVLBSKQDTSDBRAGTGTGENDKEEKKDETSSSSEANSRCQTP1KM 420  
Qy 421 KPNIEPPENWMSGASMSFRVLIGTYDNFCAIARLIGTKTCQVYEFYKSSIIAPA 480  
Db 382 KPNIEPPENWMSGASMSFRVLIGTYDNFCAIARLIGTKTCQVYEFYKSSIIAPA 441  
Qy 481 PABVDTPPRKKRKHRLMAHCKRIQLKDGSSNHVYNYQPCDHPRQPCDSCPCVIAQ 540  
Db 442 PABVDTPPRKKRKHRLMAHCKRIQLKDGSSNHVYNYQPCDHPRQPCDSCPCVIAQ 501  
Qy 541 NFCEKFCQCSSECCNRPFGCRCKAQCNTKQCPCYLAVERCDPDLCTGGAADHWSKNVS 600  
Db 502 NFCEKFCQCSSECCNRPFGCRCKAQCNTKQCPCYLAVERCDPDLCTGGAADHWSKNVS 561  
Qy 601 CKNCISIGSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISODEADRGRKYDK 660  
Db 562 CKNCISIGSGSKHLLAPSDVAGWGIPIKDPVQKNEFISEYCGEIIISODEADRGRKYDK 621  
Qy 661 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 699  
Db 622 YMCSEFLFNANDPVVDATRKGNKIRFANHSVNPNCYAKV 660

## RESULT 3

US-09-589-892B-11  
Sequence 11, Application US/09589892B  
Patent No. 6689583  
GENERAL INFORMATION:  
APPLICANT: Jenuwein, Thomas  
APPLICANT: Laible, Gotz  
APPLICANT: O'Carroll, Donal  
APPLICANT: Eisenhaber, Frank  
TITLE OF INVENTION: Chromatin-Regulator Genes  
FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/09/589, 892B  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945,988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
PRIOR FILING DATE: 1995-05-10  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 11  
LENGTH: 760  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-589-892B-11

Query Match 50.8%; Score 1934.5; DB 2; Length 760;  
Best Local Similarity 52.1%; Pred. No. 2.9e-159;  
Matches 395; Conservative 85; Mismatches 165; Indels 113; Gaps 17;

Qy 8 SEKGPVCKRKKVSEYMLRQLKFRPADYKVSFSSSRQKILERTTEIL--NOEMKORR 64  
Db 3 STKVPEKRRKVSSEIKIRQKKYKRADEIKEMIRNWDHNNVNDLYESKVMQAKP 62  
Qy 65 IQPVHILTSVSLRGTRCSVTSLDLPT--QVILPKTLNVAASVDMYSWSPLQONFMV 122  
Db 63 YDPHIV-----DCVRAVYTSYNGIIPSGQKVPICBINAVVTPIPTMYTAAPTQONFMV 115  
Qy 123 EDETVLANIPYMGDEVLDQDGTPIBELIKNYDGKVGDRPCGFINDEIFVELVNAL----- 178

Db 116 EBEYLAHNPVMDVLDKDGKFIIEILINVOGKHGDDPSPMDAIVELVHMLMSY 175  
179 -----GQYND- DDDGDGP- EBEKQKLE- 205  
Db 176 SKELEAPSTAIKTEBPLAKSGEDDGVVDVADCS PMKLETKESGDLTVKEKE 235  
Qy 206 -----HRDKSRPRK-----PSSDIFPAISMPDCKGTABELKEKYLEBQOLP 253  
Db 226 TEBPEVEDADVPKAVEVKDLPPAPPIFOAISANFPDKGTABELKEKYLEBQOLP 295  
Qy 254 GALPPECTINIDGPAKSVQREOSLHSPFTLLFCRCFKDCFLAPF--HATENTYKAKT 311  
Db 226 -ERPECTINIDGPAKSVQREOSLHSPFTLLFCRCFKDCFLAPF--HATENTYKAKT 354  
Qy 312 ETALDNKPCGPOCYOHLBGAKEPAAALTAERIKTPPKRGRRGLPNNSSRPSTPTIN 371  
Db 355 ELKPPAPPCSNCYMLIDGMEKLAADS-----KTPP----- 386  
Qy 372 VLBSKDTSDREAGETGGENDK-----EBEKQETSSSSSEANSRCQTPIMKPIY 424  
Db 387 -----IDSCNASESDNSOPSNDPNNHNSKNDGLTVNSAAVAINISIMGMNI 440  
Qy 425 BPEPVMSGAASMPVILGYYNFCAIALIGTKROYBRVKSIIADAPAD 484  
Db 441 TSTQCV-WTGAQALYVLAHYLKYKCALANMLTKROYBRVKSIIADAPAD 496  
Qy 485 VD---TPPKKKRKLMAHCRKIQLKKOGSSNNVYVOPCDHPROPDSSCPVIAON 541  
Db 497 LRQDTPPRKKKKQKRLMGLHCRKIQLKKOGSSNNVYVOPCDHPROPDSSCPVIAON 556  
Qy 542 PEEKCCOSSSECONPFGRCRCAQCTKQPCYLAVERCDPLCLTGAADHMSKNSVC 601  
Db 557 PEEKCCOSSSECONPFGRCRCAQCTKQPCYLAVERCDPLCLTGAADHMSKNSVC 615  
Qy 602 KNCSTGSKKLLAPSDVAGMGIPIYDPOKNEPISYCGEIIISODEADRGKYVYK 661  
Db 616 KNCSTGSKKLLAPSDVAGMGIPIYDPOKNEPISYCGEIIISODEADRGKYVYK 675  
Qy 662 MCSFLFNANDPVVDAIRKGNKIRFANHSVNPCTYAKV 699  
Db 675 MCSFLFNANDPVVDAIRKGNKIRFANHSVNPCTYAKV 713

## RESULT 4

US-09-699-266A-13

Sequence 13, Application US/09699266A

Patent No. 6559354

GENERAL INFORMATION:

APPLICANT: Ma, Hongchang

APPLICANT: Morakinyo, Layo O.

APPLICANT: Odell, Joan T.

APPLICANT: Orozco Jr., Emil M.

APPLICANT: Rafalekt, J. Antoni

TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS

FILE REFERENCE: Bb1164 US NA

CURRENT APPLICATION NUMBER: US/09/699,266A

PRIOR FILING DATE: 2000-10-27

PRIOR APPLICATION NUMBER: PCT/US99/08385

PRIOR FILING DATE: 1999-04-16

PRIOR APPLICATION NUMBER: 60/083,212

NUMBER OF SEQ ID NOS: 13

SOFTWARE: Microsoft Office 97

SEQ ID NO 13

LENGTH: 856

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-09-699-266A-13

Query Match 18.9% Score 719.5; DB 2; Length 856;  
Best Local Similarity 27.6%; Pred. No. 1,5e-53;  
Matches 231; Conservative 110; Mismatches 256; Indels 241; Gaps 38;

Qy 25 ELKQKRRBRADDEVKSM-----FSSNRKILTEILINQEMKORIOPIVHILTSVSLRGR 81  
Db 35 ELKQKRRBRADDEVKSM-----FSSNRKILTEILINQEMKORIOPIVHILTSVSLRGR 79  
Qy 82 ECVTSVSDLPFQVPLKTLNVAVASVPIMYSWSPLOQFMVEDFTVHNI-----P 132  
Db 80 EDNGSNMNLSSRMHPLCKLNG-----FSGVGDRLVYVPTKOVISASVYKPIARLIP 132  
Qy 133 YMGDEVLDQDGTPIELIKNTDGKVGRCGFINDLIFVELVNAAGQYND- DDDG 190  
Db 133 YTWVIFLDRNQMAE-----DQSVGRR-----QIYR-----OHGETTILCSDE 173  
Qy 131 DDPERERKQKDLERHDKSRPRPKPSDKIFPAISMPDCKGTABELK----- 241  
Db 174 BPEPEERKRRSRSG-----ESITWILQOEY-----GMGEVQDALCOLLSV 216  
Qy 242 -----EKYKELTEQOLFGALPPECTPNDGPNNAK-SVORQ-----SLHSFTLLFCRCF 290  
Db 217 DASDILERYNBLKLD-----KONTEPSSNGFGLISLEKGLAALDSFNLFCRCL 270  
Qy 291 KYDCFLAPPHATPNTYKRN-----TETALDNKPCGPOCYOHLBGAK-PA----- 336  
Db 271 VFDCLHGCQPLISASEKQPYWSDYEDRKRCSHCYQLKAVAREVETCSNFASKAE 330  
Qy 337 -----ALTAER-----IKTPKRG-----GRR----- 355  
Db 331 KASEBCKRAVSDVPHAAASGVSLOVEKTDIGINVDSSGVEQEHGIRGRVPLKD 390  
Qy 356 -GRLEPNSSRP-----STPTINVLBSKDTSDREAGETGGENDKXEBEK 400  
Db 391 SNDLPNLSNKQKTAASPTKMSFVNSVPSLD--QALDSTKDDQGTITNKVNRDEADAK 448  
Qy 401 K-----DETSSSEANSRCQTP-----IKKPNIEPVMSGAASMPVILG 445  
Db 449 EYGEPIPNDSVHDDGSSICQPHHSGNGAIIIAEMSESRP--STEMPIEDLYIKVE 506  
Qy 446 TYTDFNFCALAR--LIGTTCQVYEF-RVKSIIAPA-----PAEYD--TP 488  
Db 507 IFGRNSCLIANRLSGLTKCDVSNYMEENSVSRSTPNLLDDRTDPGNNDVVP 566  
Qy 489 PR-----KKGRKRLM-----AAHCRKIQLKKOGSSNNVYVOPCDHPROPDSSCPVY 538  
Db 567 PTRLFRKKGKTRKLTSTKAGHPSWKRIAGNGNSCKQYTPCG-CLMSGKDCPCLT 625  
Qy 539 AONPEKPCOSSSECONPFGRCR-KAQCNTPQPCYLAVERCDPLCLTGAADHMSK 597  
Db 626 NETCEKCYGCKSKCONFRGCHCAKSGCRSGCPFAAGRECDPVCANC-----W--- 677  
Qy 598 NVSCNCSI---QRG-----SKHLLAPSDVAGMGIPIYDPOKNEPISY 641  
Db 678 -VSCGDGSLGAPRGEGQCNMRLLRQOORILLGKSDVAGMGAFLNVSVKNEYLYGEY 736  
Qy 642 CGEIIISODEADRGVYDKWMSFLFNANDPVVDAIRKGNKIRFANHSVNPCTYAKV 699  
Db 737 TGEIISHHADKRGKTYDRANSSFLFINDQVLDQKRGDLKAFANHSVNPCTYAKV 794

## RESULT 5

US-09-699-266A-12

Sequence 12, Application US/09699266A

Patent No. 6559354

GENERAL INFORMATION:

APPLICANT: Ma, Hongchang

APPLICANT: Morakinyo, Layo O.

APPLICANT: Odell, Joan T.

APPLICANT: Orozco Jr., Emil M.

TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS

FILE REFERENCE: Bb1164 US NA

CURRENT APPLICATION NUMBER: US/09/699,266A

PRIOR FILING DATE: 2000-10-27

PRIOR APPLICATION NUMBER: PCT/US99/08385

PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/083,212  
PRIOR FILING DATE: 1998-04-27  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 12  
LENGTH: 902  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-699-266A-12

Query Match 17.8%; Score 679.5; DB 2; Length 902;

Best Local Similarity 25.5%; Pred. No. 4.8e-50;  
Matches 230; Conservative 107; Mismatches 269; Indels 295; Gaps 35;

9 EKGPCKWRYKSEYMRRLQKRFPRADVKSMSFNOKILERTILNOEMKQRIQPV 68  
24 ERGPA-SKEVSEVIESLKKLAADRCISIKRIDENKKNLPAITQPMR----- 71  
69 HILTSVSLRGTRBCSVTSDL-----DPTQVILPLKTLN-----AVASV 107  
72 -----SSMERGSGCKDSDLLVYKQRDSPGKSGIDSSNNRRVYEDGPASSGWWQSSV 125  
108 PIMTSWSPLOQNFWEDEVTLNITPYMGDEVLDQGTTEILIKYDGKHGDEKCGFIN 167  
126 FVKISLRPIKM-----PDIKLSPTTWFVFLDRNQRMTE-----DOSVVGRR----- 167  
168 DEIPVELVNALGQVYNDDDDDGDDPEREREKQKOLEHRODKESRPPRF----- 218  
168 -RIYYDQGEALICSDSBEALIDBEK-----RDLPLB---PEDYIIRNTLQOL 213  
219 -PSDKIFPAISMPDPKGTAEELKEKYKELTEQOLPGALPRECTENIDGPNAKSV---QR 274  
214 GLSDSVLAEIAS-FLSRST-SIKARHGLVMEK-----EVSSGDNQALSSLLNDOM 264  
275 EOSLHSFTLLFCRCFCYKDCFLH-----PHATPNTYKRTKATETALDNKPPGCP 322  
265 EGALDSFNLFCRCCLVDFCRHGCSDLIIPAKRPAWCP-----PVDEWLTGCA 315  
323 QCYQLBEAKRFAALTLAR-----RIKTPKRGGRGRGLPNN---SSRSTPTI 370  
316 NCTKLKLSGRPPGYGTIGKGTGSSDAGIKTITTKSSSKLNGKRPKTFPESBSAEMK 375  
371 NVLAS-----KOTDSD-----REAGTETGGENN----- 393  
376 CALFETSDENGLQDTNSDKVSSSPKYSGRRVGRKKNKVAERVPKTOXQKTEA 435  
394 -----DKREBKODETSS-----SEANRCQTPIMAKPV----- 423  
436 SDSDSIASGSCSPDAKHKDNEATSSQKHVSGNSGKSRKGTGPABVSNNSVKDDVPV 495  
424 -----IBPENVE-----WSGABSMFVILIGTYDNFC 452  
496 CQSNFVASELDAPGSDSLRKEEFMGETYSRGLATNKLMLRLESLEFPGVETIGAMSC 555  
453 AIAF--LIGTKTCROYVEFRV-----KESSIIAPAPADVDTPPRKQKK 495  
556 LIARILSGFSKSWFYQWTCSEMKASFPGGDGLNPOSSKFDLNGMNVNNOVRRSRHF 615  
496 HR-----LW-----AAHC--RKIQLKQDSSNHVNYIQCPDHRO-----PCDS 532  
616 LRRRGVRLKTYWKSAAVHSIRKITEKDD-----QPC---RQFPNCNCTIACGK 663  
533 SPCYIAQNFCEKFCQCSSECONRPPGCGC-KAQCNTKQCPYLAVRBCDPLCLTGAA 591  
664 ECPCLLNGTCCKCYGCPKSCNRFRGCGACASQCCPCPADRECDPVCANC--- 720  
592 DHM-----DSKIVSCKNCSIQSGSKHLLAPSDVAGNIGFIKDPVQKNEFI 638  
721 --WVIGGDSGLVPSRGDNVYECRNKILLKQOQRLVLSIDVSGGAVLKNYSVSHBYL 778  
639 SBYCEIISQDAEDRRGKYYDKYMGSLFNLANDPVVDATRKGNKTRFANHSVNPYCYAK 698

Db 779 GEYTGELISHKADKRGKTYDRENCSPFLNLDQFVLDAYRKGDKLKRFANHSPEPNCYAK 838  
QY 699 V 699  
Db 839 V 839

RESULT 6  
US-09-699-266A-9  
Sequence 9, Application US/09699266A  
Patent No. 6559354

GENERAL INFORMATION:  
APPLICANT: Ma, Hongchang  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Odell, Joan T.  
APPLICANT: Orozco Jr., Emil M.  
APPLICANT: Rafaleki, J. Antoni

TITLE OR INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS  
FILE REFERENCE: B1164 US NA  
CURRENT FILING DATE: US/09/699,266A  
PRIOR APPLICATION NUMBER: PCT/US99/08385  
PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/083,212  
PRIOR FILING DATE: 1998-04-27  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 9  
LENGTH: 921  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-699-266A-9

Query Match 17.5%; Score 665; DB 2; Length 921;

Best Local Similarity 25.5%; Pred. No. 9e-49;  
Matches 230; Conservative 116; Mismatches 248; Indels 308; Gaps 41;

16 RRYKSEYMRRLQKRFPRADVKSMSFNOKILERTILNOEMKQRIQPVHILSVS 75  
49 KKRITAD--RLTYK--NRIGENKTNISSTYQ-----RTYNLS---KNNQISTSKGTDSAS 97  
76 SLRGTR-----CSVTSDDF-----PTQVI--PLKTL 101  
98 NLTTRQDDALCTLHS-IDITPVDDKGTFPDSEPFSSNNMFGANLGPKNAILIRPIK-L 155  
102 NAVASVPIMYSWSPLOQNFWEDEVTLNITPYMGDEVLDQGTTEILIKYDGKHGDEKCGFIN 160  
156 PEVPLPPTTWIFLDRNQRMTEQSVLGR-----RRIYYD 191  
161 RECGFINDEIVELVNALGQVYNDDDDDGDDPEREREKQKOLEHRODKESRPPRFK 217  
192 TSCG---EALI-----CSDSBEALIEDEBEKKEFGHSBDHITRMTOEGC--- 233  
218 FPSDKIFPAISMPDPKGTAEELKEKYKELTEQOLPGALPP-----ECTPNIDG 266  
224 -MSDAVLQTLARHM--EPAADDIRKRYIILGKTKYSCCKGTEHNVVBDLYCDXKDLDA 290  
267 PNAKSVQEOSLHSFTLLFCR-----RCFKYDCFLH-----PHATPNTYKRTKETA 314  
221 -----ALDSNDNLCFRPREGRCLVDFCRHGCSDGLVFP-----TEKQPMAGV 334  
315 LDNRCGPOCYOHEGATEFPAALTAER---IKTPP-----KRGGRRR-----GRL 358  
335 DSVVCGIHCML---ASEPDAAAGADHMLFVDEPFTSSDNNVNOQPSNRRKNGSSGRK 391  
359 PNNRSRPTPTINTL-EKQND----- 379  
392 TKSQSSESSSTARVISESSDSDEVAHISNKSFOHSPSPKVKIGPYGKIRKITTNRARIARI 451  
360 -----SD-----REAGTETGGENNDE-----EERKQDET 404  
452 LMSVYKQGERMASSDSNVSGLYLLARDMKLSDPTNGKMKELLVSSQSSPSTRSSKCKST 511

[illegible]

RESULT 7  
 US-09-177-249-2  
 ; Sequence 2, Application US/09177249  
 ; Patent No. 6229064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Robert L.  
 ; APPLICANT: Ohad, Nir  
 ; APPLICANT: Klyosue, Tomohiro  
 ; APPLICANT: Yadegari, Ramin  
 ; APPLICANT: Margossian, Linda  
 ; APPLICANT: Harada, John  
 ; APPLICANT: Goldberg, Robert B.  
 ; APPLICANT: The Regents of the University of California  
 ; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
 ; TITLE OF INVENTION: Development in Plants  
 ; FILE REFERENCE: 023070-086120US  
 ; CURRENT APPLICATION NUMBER: US/09/177,249  
 ; CURRENT FILING DATE: 1998-10-22  
 ; EARLIER APPLICATION NUMBER: US 09/071,838  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 324  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 689  
 ; TYPE: prt  
 ; ORGANISM: Arabidopsis sp.  
 ; US-09-177-249-2

	Query Match	16.1%;	Score 612;	DB 2;	Length 689;	
	Best Local Similarity	25.4%;	Pred. No. 2.4e-44;			
	Matches 162;	Conservative 100;	Mismatches 197;	Indels 180;	Gaps 21;	
Cy	120 FMEVEEYLAHNIPLYMGDEVLDDGGFPI-----BELIKNTDGVKVGREGCGFINDEIFVELV	175				
Db	114 FLDEEDVPLPEVKELPIYEKLPRSTITWVTKSQMAESD-STIGRGQIYLNGE-----	166				
Cy	176 NALGQYNDDDDDDDGDPPER-REEKKOLLEDRHDKESRPPKFPSPDXIIFEAISMFDPK	234				
Db	167 -ALETSSSEDEDEDEBEERIKKECKERSSED-----VDREITWVGQDY	208				
Cy	235 G-----TAEELKEKYKRLITBQLPGALPPECTPINIDGPNAKSVGRBSL	278				
Db	209 GLDDLIVRRALAKYLEVDVSILREYNELKLN-----DTAGAGA--SDLTLS	253				

QY	279	HSFHLF-----	CRGCFKDCGLHPHAAENNYAKRANSTAL----	DNKPGPGCYGH	327
Db	254	KTITAFDPDRRRCRCMIFDCMHMEKY	-EPERSSDSKSLPREDRQPCSEKCYLK	312	
QY	328	LEGAEFAALTAERIKTPPKRPGRRRGRLFNSSRPSTPINVLESKOTSDREAGTE	387		
Db	313	VRSVYTE-----	ADHV-----	MDNDS-----	328
QY	386	TGGRNDBEBBKKDQETSSSEBANSRCQPIIMKNIIEBPENVEWGAEMFRVLIGTY	447		
Db	329	-----	ISNKIVSDPNNTMTTPEVKDIYLGKIBIF	358	
QY	448	YDNFCALR--LIGTKTCROYVEPRVKESSI-----	IAPAPADVD	486	
Db	359	GRNSCDVALNLTIRGLKTCLEIYNYVRBDQCTMSLDLNTKTTGRHNQVTKVSRKKSRSV-	417		
QY	487	TPPRKKRKHFLMAAHCKIQLKDGSSNNHYNYNPPCHPRPQPCSSGCVCAVQNFCEPE	546		
Db	418	----	RKSRRLKRYAAPPALMKTTTSGAKFYKHYPCT-CKSKCGQOQCCLTHENCCEKY	472	
QY	547	CGSSBCCNRPFGACCK-AQCNTRKQCPCYLAVRECDPDLC-----	LTCGADAHMDSK-ANYS	600	
Db	473	CGCSBDCNNRRGGCNCALGQCTNRCQCPFPAANRECDPDLCRSCPLSCSDGTIGETVQIQ	532		
QY	601	CKNCSIORGSKRHLLLAPSDVAGWGIPTKDPVQXKNEFISXYGAILISODADRGKRVYDK	660		
Db	533	CKNNQFLLQTNKILLIGKSDVHWGMAFTMWSLAKNREYLGAEYGGELITHDEANBRGRIBDR	592		
QY	661	YMGCFPLTYLNDPVDVATRKGNKIRFPAHSHVNPNCYAYV	699		
Db	593	IGSSYLFPLTINDQLETDARKKNEBFGYLNHSAFNPCTALD	631		

```

RESULT 8
US-09-061-769A-2
; Sequence 2, Application US/09061769A
; Patent No. 6239327
; GENERAL INFORMATION:
; APPLICANT: Cold Spring Harbor Labs
; TITLE OF INVENTION: No. 6239327a1 Seed Specific Polycomb Group Gene and Methods of
; TITLE OF INVENTION: use for Same
; FILE REFERENCE: mea genes
; CURRENT APPLICATION NUMBER: US/09/061,769A
; CURRENT FILING DATE: 1998-04-16
; PRIOR APPLICATION NUMBER: 09/061,769
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEAWT
US-09-061-769A-2

```

```

~ Query Match      16.1%; Score 612; DB 2; Length 689;
Best Local Similarity 25.4%; Pred. No. 2.4e-44;
Matches 162; Conservative 100; Mismatches 197; Indels 180; Gaps 21

QY 120 FVVEDEVLANIPLYMGDEVLDDGDTFI-----BELIKNYDKGVHGRCEGFINDIEFVELV 175
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 114 FLDEVDPLPLPSVKLPIYEKLPRSLTWPTTKSSQLMAESD-SVIGRQRQIYYLNGR----- 166

QY 176 NALGQYNDDDDDDDDGGDPPEE-RERKQKDLBHRDXKSRPRKKFSDDKIFELISSMFDPX 234
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 167 -ALELSSEBDBEDBEDEBKKEKCEPSSD-----VDRFTWVGQDY 208

QY 235 G-----TAEELKERYKYLTEQQLPGLAPPECTPNIDGPNAKSVOREOSL 278
   ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 209 GLDDLVVRRAAKTLKVLDVSDILERYNELKLN-----DETAGA--SDLTS 253

QY 279 HSFHTLF-----CRRCFYDCFLAHFHATPNTYKRKNETIAL--DNKCPGQCIOH 327

```

Db 254 KITTAFOADPADRRHCHRCMIFFDCHMEKY-BEPSRSSSEKSSIFBEDRQPCSEHCYK 312  
Qy 328 LEGAKEPAALATARIKTPPRKPGRRGRGLPNNSRSTPTINVLKSDTSDREAGTE 387  
Db 313 VRSYTE-----ADHV-----MDNDS----- 328  
Qy 388 TGGENDKEBEKDEKOTSSSEANSRCQTPYKKNPIEPPEVWMSGABSMFVILGTY 447  
Db 329 -----ADHV-----MDNDS----- 328  
Qy 448 YDNFCALR--LIGTKCROYERFRVKESS-----IAPAPAEVD 486  
Db 359 GRNSCDVALNIRGLKCLRIYNYWRBODCTMSLDINKTTOHNOVTKKYSRKSRSV- 417  
Qy 487 TPRKRRKRLMAHCKRIQLKKDGSNNHYNYOPCDHPRQPCDSSCPVIAONFCEKF 546  
Db 418 ---RKSRLKRYAPYPPALKKTTSGBAKFKYHTPCT-CKSKGQCCPCLTHENCCEKY 472  
Qy 547 CQSSSECONRPPGCRCK-AQCNTKQCPCYLAVRBCDPLC---LTCGADHMDSK-NVS 600  
Db 473 CGSKDCNNRRGGGNCALIGQCTNRQCPCFANRBCDPLCRSCPLSGDGLTIGTPVQIQ 532  
Qy 601 CKNSIQGSKGHLILAPSDVAGWGIPIKDPVQKNRPISEYCGEIIISQDEADRGRKYDK 660  
Db 533 CKNQFLQITNKKILIGKSDVHGWAFTWDSLKNRBYGELYGBLITHDENRGRJEDR 592  
Qy 661 YMCSPFLNANDFVVDATRKGNKIRIPANHSVNPVCYAK 699  
Db 593 IGSSTLFTLNDQLEIDARRKGNRKFPLNHSARPCYAKL 631

RESULT 9  
US-09-812-283-2  
Sequence 2, Application US/09812283  
Patent No. 6828477  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Ohad, Nir  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadegari, Ramlin  
APPLICANT: Margosian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
TITLE OF INVENTION: Development in Plants  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/812,283  
CURRENT FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 09/177,249  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: US 09/071,838  
PRIOR FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 689  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-09-812-283-2

Query Match 16.1%; Score 612; DB 2; Length 689;  
Best Local Similarity 25.4%; Pred. No. 2.4e-44;  
Matches 162; Conservative 100; Mismatches 197; Indels 180; Gaps 21;  
Qy 120 FVNEDEVYAHNIPYMGDEVLDQDGTFI---BELIKNDGKVGHRCGFINDEIFVELV 175  
Db 114 FVDEVPILPSVYLPIVERLPRISITWTFKSSQLMAESD-SYIGKQIYIYNGE----- 166  
Qy 176 NALGQYNDDDDDDDGGDPRR-RBEKQKQJEDHRDDDESPPKPFSDKIFEAISMFDPK 234  
Db 167 -ALELSEDEDEDEDEBEIKKEKCEFSDE-----VDRFIWVGQDY 208

Qy 235 G-----TABELKRYKELTEQOLPGALPPECPTNIDGPNKSVQREOSL 278  
Db 209 GLDDLVRRALAKYLEVDVSDILERYNELTKN-----DGTAGEA--SDUTS 253  
Qy 279 HSHFTLF-----GRCFKQDCPLHPFATNTYRKNTETAL---DNKPCGPOCYQH 327  
Db 254 KITTAFOADPADRRHCHRCMIFFDCHMEKY-BEPSRSSSEKSSIFBEDRQPCSEHCYK 312  
Qy 328 LEGAKEPAALATARIKTPPRKPGRRGRGLPNNSRSTPTINVLKSDTSDREAGTE 387  
Db 313 VRSYTE-----ADHV-----MDNDS----- 328  
Qy 388 TGGENDKEBEKDEKOTSSSEANSRCQTPYKKNPIEPPEVWMSGABSMFVILGTY 447  
Db 329 -----ADHV-----MDNDS----- 328  
Qy 448 YDNFCALR--LIGTKCROYERFRVKESS-----IAPAPAEVD 486  
Db 359 GRNSCDVALNIRGLKCLRIYNYWRBODCTMSLDINKTTOHNOVTKKYSRKSRSV- 417  
Qy 487 TPRKRRKRLMAHCKRIQLKKDGSNNHYNYOPCDHPRQPCDSSCPVIAONFCEKF 546  
Db 418 ---RKSRLKRYAPYPPALKKTTSGBAKFKYHTPCT-CKSKGQCCPCLTHENCCEKY 472  
Qy 547 CQSSSECONRPPGCRCK-AQCNTKQCPCYLAVRBCDPLC---LTCGADHMDSK-NVS 600  
Db 473 CGSKDCNNRRGGGNCALIGQCTNRQCPCFANRBCDPLCRSCPLSGDGLTIGTPVQIQ 532  
Qy 601 CKNSIQGSKGHLILAPSDVAGWGIPIKDPVQKNRPISEYCGEIIISQDEADRGRKYDK 660  
Db 533 CKNQFLQITNKKILIGKSDVHGWAFTWDSLKNRBYGELYGBLITHDENRGRJEDR 592  
Qy 661 YMCSPFLNANDFVVDATRKGNKIRIPANHSVNPVCYAK 699  
Db 593 IGSSTLFTLNDQLEIDARRKGNRKFPLNHSARPCYAKL 631

RESULT 10  
US-10-176-884-14  
Sequence 14, Application US/10176884  
Patent No. 6906244  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert  
APPLICANT: Kinoshita, Tetsu  
APPLICANT: Yadegari, Ramlin  
APPLICANT: Gehring, Mary  
APPLICANT: Okamuro, Jack  
APPLICANT: Dang, Van-Dinh  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
FILE REFERENCE: 023070-116710US  
CURRENT APPLICATION NUMBER: US/10/176,884  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/300,506  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 689  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: MEDEA (MEA)  
US-10-176-884-14

Query Match 16.1%; Score 612; DB 2; Length 689;  
Best Local Similarity 25.4%; Pred. No. 2.4e-44;  
Matches 162; Conservative 100; Mismatches 197; Indels 180; Gaps 21;  
Qy 120 FVNEDEVYAHNIPYMGDEVLDQDGTFI---BELIKNDGKVGHRCGFINDEIFVELV 175





```

; APPLICANT: Kinoshita, Tetsu
; APPLICANT: Yadegari, Ramin
; APPLICANT: Gehring, Mary
; APPLICANT: Okamoto, Jack
; APPLICANT: Dang, Van-Dinh
; APPLICANT: The Regents of the University of California
; APPLICANT: Ceres, Inc.
; TITLE OF INVENTION: Compositions and Methods for Modulating Plant
; TITLE OF INVENTION: Development
; FILE REFERENCE: 023070-116710US
; CURRENT APPLICATION NUMBER: US/10/176,884
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/300,506
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEA-15G-ESAI
; OTHER INFORMATION: chimeric polypeptide encoded by transgene plasmid
; OTHER INFORMATION: pCRS304-50AS MEA-15G-ESAI
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (906)
; OTHER INFORMATION: Xaa = Gly or Glu
; US-10-176-884-15

Query Match      16.1%; Score 612; DB 2; Length 932;
Best Local Similarity 25.4%; Pred. No. 3.7e-44;
Matches 162; Conservative 100; Mismatches 197; Indels 180; Gaps 21;

QY 120 FAVEBETLHNPVMDVLDGDTPI---BELKNDGKNGHRECFINDLIFVELV 175
DB 114 FLDEVPPLPSVYKPIVEKLPSTWVFTKSSQLMAESD-SYIGRQIYILNGE----- 166
QY 176 NALGOYNDDDDDGDDGDPPE-REBKQKOLEDRDDKESRPPKPSDKIFBAISSMPDK 234
DB 167 -ALBESSERDEDEDEDEIKKCKECPSED-----VDRIWTVGQDY 208
QY 225 G-----TABELKEKYKEITLQOLPQALPEECTPNIDGNNAKSVORBSL 278
DB 209 GLDDLVRRALAKYLEVDVSDILERYNELKLN-----DGTAGBA--SDLTLS 253
QY 279 HSFHLP-----CRGFKYDCPLPHATPNYTKRNTETAL---DNKCGPQCYH 327
DB 254 KITTAFODFADRRHCRRCMIFDCMHKRY-BPESRSSDKSLPEDEDRQPCSEHCYIK 312
QY 328 LEGAEPFAALTAERIKTPPKRPGRRGRRLPNNSRPSPTPTINVLSEKOTSDREAGTE 387
DB 313 VASVTE-----ADHY-----MUNDN----- 328
QY 388 TGENNDKEEBEKDETSSEANSRCQTPPIKPKPIEBEPENVSGAASMFVILGTY 447
DB 329 -----ISNKIIVSDPNTMTMTPEVKDLYLKGIKIF 358
QY 448 YNFCALAR--LIGTKCQVYEFVYKESI-----IAPAEVDV 486
DB 359 GNSDCDVANLIRGKTKLEIYNMEBOQCTMSIDLNTKTRHNOVTKVSRKSRSTV- 417
QY 487 TPRKKKGRRLMAAHCRIOLKKQSSNHVYNYOPCDHROPDSSCPVLAQNFCEKF 546
DB 418 ----RKSRIRKRYARPPALAKTTTSGEALFYKHTYPTCT-CKKCGQCCCLTHENCERY 472
QY 547 CQCSSECONRPPGCRCK-AQCNTPQCPCYLAVERCDPDLCT---LTGAADHMSK-NVS 600
DB 473 CGCSXDCNNRFGGCAICGCTNBOCPFAANRECPDLCRSCLSGDGTGAEFPVQIO 532
QY 601 CKNCSIQRSKSHLLAPSDVAGWGIPIQDPVQKNEFISFYGGEIISQEAARGRGVYK 660
DB 533 CKNMOPFLQTNKKILIGKSDVHGWGAFTWDSLKQVEYILBYGELLTHDEANERGRIEDR 592
```

```

QY 661 YMCSEFLNANDFVVDATRKGNKIRFANHSVNPYCAKV 699
DB 593 IGSSYLFNLNDQLEIDARKKGNFEPFLNHSARPNYCAKL 631

RESULT 13
US-09-699-266A-11
; Sequence 11, Application US/09699266A
; Patent No. 6559354
; GENERAL INFORMATION:
; APPLICANT: Ma, Hongchang
; APPLICANT: Morakinyo, Layo O.
; APPLICANT: Odell, Joan T.
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Rafalekt, J. Antoni
; TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS
; FILE REFERENCE: B01164 US NA
; CURRENT APPLICATION NUMBER: US/09/699,266A
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/US99/08385
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/083,212
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Triticum aestivum
; US-09-699-266A-11

Query Match      12.4%; Score 470.5; DB 2; Length 359;
Best Local Similarity 36.3%; Pred. No. 1.9e-32;
Matches 109; Conservative 39; Mismatches 119; Indels 33; Gaps 7;

QY 432 WSGAASFPRLIGTYNFCALAR--LIGTKCQV--YEFVYKESIAPAEVDV 487
DB 3 WSTLERDLYLKGIEIFGNISCLIVNLLGKLTCEVASVYNGAANMSKISIGDTET 62
QY 488 PP-----RKKRKRRLMAAHCRIOLKKQSSNHVYNYOPCDHROP 529
DB 63 EONTMEQGMVVRITKVCRRGRTRKHKYPSKAAGPAIRAKYGDGQCDROGTPG-CQEM 121
QY 530 CDSGCPVIAONFCEKFCQCSSECONRPPGCRCK-AQCNTPQCPCYLAVERCDPDLCTC 588
DB 122 CNKNCPCVENGTCCEKYCGCSKSCNRRFRGCHCAKISAEAGNAHYLTPSGNVIMFCHNC 181
QY 589 -----GAADHMSK-NV--SCNCSIQRSKSHLLAPSDVAGWGIPIQDPVQKNEIS 639
DB 182 WGEIWSGSPRRATKKEVWVTOCGNMKLLKQOORILLKSDVAGAGATIKPVHKNIDVLG 241
QY 640 FYCGEIIISODEADRRGRKYDKYKMSFLPNLNDFVVDATRKGNKIRFANHSVNPYCAKV 699
DB 242 EYTGELISHKADRKGLIYDRANSFLFDLNDQFLVDYLRKGDGLKFPANHSNPYCAKV 301

RESULT 14
US-10-200-012-39
; Sequence 39, Application US/10200012
; Patent No. 6955905
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR/SET- Domain Containing Nucleic Acids,
; TITLE OF INVENTION: Polypeptides, Antibodies and Methods of Use
; FILE REFERENCE: P-117 5301
; CURRENT APPLICATION NUMBER: US/10/200,012
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 09/910,478
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
```

LENGTH: 114  
TYPE: PRT  
ORGANISM: Human  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,  
LOCATION: 46, 47, 48, 49, 50, 51  
OTHER INFORMATION: synthetic peptide  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-200-012-39

Query Match 7.7%; Score 293; DB 2; Length 114;  
Best Local Similarity 73.4%; Pred. No. 1e-17;  
Matches 58; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 621 VAGWGIFIDPVQKNEFISYCGEIIISODEADRGKVDKMCSEFLNINDFVVDATRK 680  
DB 1 VAGWGIFIDPVQKNEFISYCGEIIISODEXXXXXXXXXXXXXXXXXXXXXDPVVDATRK 60

QY 681 GKIRPANHSVNPNCYAKV 699  
DB 61 GKIRPANHSVNPNCYAKV 79

RESULT 15  
US-10-176-884-19

Sequence 19, Application US/10176884  
Patent No. 6906244  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert  
APPLICANT: Kinoshita, Tetsu  
APPLICANT: Yadegari, Ramin  
APPLICANT: Gehring, Mary  
APPLICANT: Okamura, Jack  
APPLICANT: Dang, Van-Dinh  
APPLICANT: The Regents of the University of California  
APPLICANT: Ceres, Inc.  
TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
TITLE OF INVENTION: Development  
FILE REFERENCE: 023070-116710US  
CURRENT APPLICATION NUMBER: US/10/176,884  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/300,506  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 19  
LENGTH: 115  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:exemplary SET  
OTHER INFORMATION: domain  
US-10-176-884-19

Query Match 6.6%; Score 251; DB 2; Length 115;  
Best Local Similarity 53.1%; Pred. No. 4.6e-14;  
Matches 43; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

QY 619 SDVAGWGIFIDPVQKNEFISYCGEIIISODEADRGKVDKMCSEFLNINDFVVDAT 678  
DB 1 SDVAGWGATWTWDLKNEFISYCGEIIISODEADRGKVDKMCSEFLNINDFVVDAT 60

QY 679 RKGNKIRPANHSVNPNCYAKV 699  
DB 61 RKGNKIRPANHSVNPNCYAKV 81

Search completed: August 11, 2006, 22:30:23  
Job time : 37.8291 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:45:12 ; Search time 118,999 Seconds  
(without alignments)  
2720.917 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGQTKSEKBPVCWRKRVK.....KKNKIRFANHNVNPKYAKV 699

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	746	5	US-10-773-302-2
2	3802	99.8	746	4	US-10-153-668-232
3	3802	99.8	746	4	US-10-231-778-225
4	3802	99.8	746	4	US-10-295-027-710
5	3750	98.5	746	4	US-10-231-778-226
6	3569.5	93.7	707	4	US-10-104-047-3162
7	3569.5	93.7	707	6	US-11-072-512-3162
8	2602	68.3	517	4	US-10-153-668-354
9	2212.5	58.1	723	4	US-10-115-482-56
10	1953.5	51.3	760	6	US-11-097-143-14073
11	1952.5	51.3	760	4	US-10-231-778-224
12	1934.5	50.8	760	5	US-10-773-302-11
13	1208.5	31.7	242	4	US-10-424-599-219717
14	719.5	18.9	856	4	US-10-231-778-221
15	719.5	18.9	880	4	US-10-425-114-72833
16	717	18.8	895	5	US-10-739-930-7656
17	699	18.4	885	4	US-10-424-599-174013
18	698	18.3	893	6	US-09-906-453-4
19	698	18.3	893	6	US-11-230-145-4
20	693	18.2	891	4	US-10-310-154-599
21	677	17.8	931	4	US-10-425-115-281072
22	677	17.8	931	4	US-10-437-963-146048
23	673	17.7	895	4	US-10-437-963-146048
24	667.5	17.5	1015	4	US-10-437-963-105693
25	666.5	17.5	902	4	US-10-231-778-222
26	666	17.5	699	4	US-10-437-963-146070
27	663	17.4	931	3	US-09-906-453-2

28	663	17.4	931	6	US-11-230-145-2	Sequence 2, Appl1
29	658	17.3	985	5	US-10-732-923-15015	Sequence 15015, A
30	657.5	17.3	958	4	US-10-425-115-281076	Sequence 281076, A
31	619	16.3	509	4	US-10-425-114-57456	Sequence 57456, A
32	612	16.1	689	3	US-09-071-838-2	Sequence 2, Appl1
33	612	16.1	689	4	US-10-213-512-2	Sequence 2, Appl1
34	612	16.1	689	4	US-10-176-884-14	Sequence 14, Appl1
35	612	16.1	689	4	US-10-231-778-1	Sequence 1, Appl1
36	612	16.1	689	4	US-10-177-478-5	Sequence 5, Appl1
37	612	16.1	689	6	US-11-051-456-14	Sequence 14, Appl1
38	612	16.1	932	4	US-10-176-884-15	Sequence 15, Appl1
39	612	16.1	932	4	US-10-177-478-21	Sequence 21, Appl1
40	612	16.1	932	6	US-11-051-456-15	Sequence 15, Appl1
41	607.5	16.0	463	4	US-10-425-114-57452	Sequence 57452, A
42	601	15.8	417	4	US-10-425-114-57969	Sequence 57969, A
43	582.5	15.3	536	4	US-10-425-114-69287	Sequence 69287, A
44	569.5	15.0	346	4	US-10-425-114-57460	Sequence 57460, A
45	547	14.4	294	4	US-10-425-114-69819	Sequence 69819, A

## ALIGNMENTS

RESULT 1  
US-10-773-302-2

Sequence 2, Application US/10773302  
Publication No. US20050089880A1

GENERAL INFORMATION:

APPLICANT: Jenwein, Thomas

APPLICANT: Labbe, Gotez

APPLICANT: O'Carroll, Donal

APPLICANT: Risenhaber, Frank

APPLICANT: Rea, Stephen

TITLE OF INVENTION: Chromatin-Regulator Genes

FILE REFERENCE: 0652.1670001

CURRENT APPLICATION NUMBER: US/10/773,302

CURRENT FILING DATE: 2004-02-09

PRIOR APPLICATION NUMBER: US/09/589,892

PRIOR FILING DATE: 2000-06-09

PRIOR APPLICATION NUMBER: US 08/945,988

PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: PCT/EP96/01818

PRIOR FILING DATE: 1996-05-02

PRIOR APPLICATION NUMBER: DE 195 16 776.7

PRIOR FILING DATE: 1995-05-10

NUMBER OF SEQ ID NOS: 21

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2

LENGTH: 746

TYPE: PRT

ORGANISM: Homo sapiens

US-10-773-302-2

Query Match

Best Local Similarity 100.0%; Score 3808; DB 5; Length 746;

Matches 699; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGQTKSEKBPVCWRKRVKSEYMLRQLKFRADRYKSMSSNRQKILERTETLNQEW	60
DB	1	MGQTKSEKBPVCWRKRVKSEYMLRQLKFRADRYKSMSSNRQKILERTETLNQEW	60
QY	61	KORRIQPVHILTSVSLGTRBCSVTSLDPTPTVYIPLKTTANAVASVPIMYSWSPLOQNF	120
DB	61	KORRIQPVHILTSVSLGTRBCSVTSLDPTPTVYIPLKTTANAVASVPIMYSWSPLOQNF	120
QY	121	MVEBETVLANIPYMGDEVLDQDGTPIBELINNYGKXGDEBCGFINDEIVELVNAAGQ	180
DB	121	MVEBETVLANIPYMGDEVLDQDGTPIBELINNYGKXGDEBCGFINDEIVELVNAAGQ	180
QY	181	YNDDDDDDDGDDPSEBSEKQKOLDEHHRDDEKSRPPRKPSDKIFEAISWMPDQGTAEEL	240
DB	181	YNDDDDDDDGDDPSEBSEKQKOLDEHHRDDEKSRPPRKPSDKIFEAISWMPDQGTAEEL	240

QY 241 KKKYKLTBOQLPGALPPECTPNIDGPNKSVQREOSLHSHFTLFCRRCFKYDCFLHFFH 300  
DB 241 KKKYKLTBOQLPGALPPECTPNIDGPNKSVQREOSLHSHFTLFCRRCFKYDCFLHFFH 300  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPRPGRRGRGLPN 360  
DB 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPRPGRRGRGLPN 360  
QY 361 NSSRPTFTINVLBSKOTDSDREAGTETGGENNDKEEBKKDETSSSEANSRCOTPIKM 420  
DB 361 NSSRPTFTINVLBSKOTDSDREAGTETGGENNDKEEBKKDETSSSEANSRCOTPIKM 420  
QY 421 KPNIEPENWMSGABASMFVLIIGTYDNFCALRIIGTKTCQVYEFVYKESSTIAPA 480  
DB 421 KPNIEPENWMSGABASMFVLIIGTYDNFCALRIIGTKTCQVYEFVYKESSTIAPA 480  
QY 481 PAEDVDTPPRKKRKHRLMAHCKRIOLKDGSSNNHYNYOPCDHPROPDSSCPVIAQ 540  
DB 481 PAEDVDTPPRKKRKHRLMAHCKRIOLKDGSSNNHYNYOPCDHPROPDSSCPVIAQ 540  
QY 541 NPCEKFCQSSSECONRPFPGCRCKAQCNTKQCPCTLAVERCDPDLCTGGAADHMSKNVS 600  
DB 541 NPCEKFCQSSSECONRPFPGCRCKAQCNTKQCPCTLAVERCDPDLCTGGAADHMSKNVS 600  
QY 601 CKNCISIQSGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGIISQDEADRGKXYDK 660  
DB 601 CKNCISIQSGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGIISQDEADRGKXYDK 660  
QY 661 YMCSPFLFNLDNFVVDATRKGNKIRFANHSVNPNCYAKV 699  
DB 661 YMCSPFLFNLDNFVVDATRKGNKIRFANHSVNPNCYAKV 699

## RESULT 2

US-10-153-668-232  
/ Sequence 232, Application US/10153668  
/ Publication No. US20030092616A1  
/ GENERAL INFORMATION:  
/ APPLICANT: HONDA, Goichi  
/ APPLICANT: MATSUDA, Akio  
/ APPLICANT: MURAMATSU, Shuji  
/ APPLICANT: ISHIZAWA, Kenya  
/ TITLE OF INVENTION: STAT6 Activating Gene  
/ FILE REFERENCE: 1254-0207P  
/ CURRENT APPLICATION NUMBER: US/10/153, 668  
/ PRIOR FILING DATE: 2002-05-24  
/ PRIOR APPLICATION NUMBER: US 60/293, 172  
/ PRIOR FILING DATE: 2001-05-25  
/ PRIOR APPLICATION NUMBER: US 60/316, 031  
/ PRIOR FILING DATE: 2001-08-31  
/ PRIOR APPLICATION NUMBER: US 60/328, 403  
/ PRIOR FILING DATE: 2001-10-12  
/ PRIOR APPLICATION NUMBER: JP 2001-157043  
/ PRIOR FILING DATE: 2001-05-25  
/ PRIOR APPLICATION NUMBER: JP 2001-260681  
/ PRIOR FILING DATE: 2001-08-30  
/ PRIOR APPLICATION NUMBER: JP 2001-313175  
/ PRIOR FILING DATE: 2001-10-10  
/ NUMBER OF SEQ ID NOS: 488  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 232  
/ LENGTH: 746  
/ TYPE: PRM  
/ ORGANISM: Homo sapiens  
US-10-153-668-232

Query Match 99.8%; Score 3802; DB 4; Length 746;  
Best Local Similarity 99.9%; Pred. No. 2, 2e-265;  
Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCGTGGKSKGKPYCWRKRVKYSYEMRLQKPRPRADDEVSMSSNOKLTETELINQW 60  
DB 1 MCGTGGKSKGKPYCWRKRVKYSYEMRLQKPRPRADDEVSMSSNOKLTETELINQW 60

QY 61 KQRIQVHILITSSVSLRGTRBGSVTSDDLDPPTQVILPLKTLNANAVSPIMYSNPLOONF 120  
DB 61 KQRIQVHILITSSVSLRGTRBGSVTSDDLDPPTQVILPLKTLNANAVSPIMYSNPLOONF 120  
QY 121 MVEDETVLHNPYMGDGVLLQDDGFIEELIKNDYGVKAGDREGCFINDEIFVELVNLALGQ 180  
DB 121 MVEDETVLHNPYMGDGVLLQDDGFIEELIKNDYGVKAGDREGCFINDEIFVELVNLALGQ 180  
QY 181 YNDDDDDDGDDPPEEREKQDLEHDDKESRPPKRFPSDKIFEALISSMPDXTGABEL 240  
DB 181 YNDDDDDDGDDPPEEREKQDLEHDDKESRPPKRFPSDKIFEALISSMPDXTGABEL 240  
QY 241 KKKYKLTBOQLPGALPPECTPNIDGPNKSVQREOSLHSHFTLFCRRCFKYDCFLHFFH 300  
DB 241 KKKYKLTBOQLPGALPPECTPNIDGPNKSVQREOSLHSHFTLFCRRCFKYDCFLHFFH 300  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPRPGRRGRGLPN 360  
DB 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPRPGRRGRGLPN 360  
QY 361 NSSRPTFTINVLBSKOTDSDREAGTETGGENNDKEEBKKDETSSSEANSRCOTPIKM 420  
DB 361 NSSRPTFTINVLBSKOTDSDREAGTETGGENNDKEEBKKDETSSSEANSRCOTPIKM 420  
QY 421 KPNIEPENWMSGABASMFVLIIGTYDNFCALRIIGTKTCQVYEFVYKESSTIAPA 480  
DB 421 KPNIEPENWMSGABASMFVLIIGTYDNFCALRIIGTKTCQVYEFVYKESSTIAPA 480  
QY 481 PAEDVDTPPRKKRKHRLMAHCKRIOLKDGSSNNHYNYOPCDHPROPDSSCPVIAQ 540  
DB 481 PAEDVDTPPRKKRKHRLMAHCKRIOLKDGSSNNHYNYOPCDHPROPDSSCPVIAQ 540  
QY 541 NPCEKFCQSSSECONRPFPGCRCKAQCNTKQCPCTLAVERCDPDLCTGGAADHMSKNVS 600  
DB 541 NPCEKFCQSSSECONRPFPGCRCKAQCNTKQCPCTLAVERCDPDLCTGGAADHMSKNVS 600  
QY 601 CKNCISIQSGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGIISQDEADRGKXYDK 660  
DB 601 CKNCISIQSGSKKHLILAPSDVAGWGIPIKDPVQKNEPISYCGIISQDEADRGKXYDK 660  
QY 661 YMCSPFLFNLDNFVVDATRKGNKIRFANHSVNPNCYAKV 699  
DB 661 YMCSPFLFNLDNFVVDATRKGNKIRFANHSVNPNCYAKV 699

## RESULT 3

US-10-231-778-225  
/ Sequence 225, Application US/10231778  
/ Publication No. US20030126647A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Bildeau, Pierre  
/ APPLICANT: Chaudhury, Abdul M.  
/ APPLICANT: Dennis, Elizabeth S.  
/ APPLICANT: Koltunow, Anna M.G.  
/ APPLICANT: Luo, Ming  
/ APPLICANT: Peacock, William J.  
/ TITLE OF INVENTION: Method for inducing seed development by down-regulating  
/ FILE REFERENCE: 72-98A  
/ CURRENT APPLICATION NUMBER: US/10/231, 778  
/ PRIOR FILING DATE: 2002-11-08  
/ PRIOR APPLICATION NUMBER: 09/398, 237  
/ PRIOR FILING DATE: 1999-09-20  
/ PRIOR APPLICATION NUMBER: 60/101, 184  
/ PRIOR FILING DATE: 1998-09-21  
/ PRIOR APPLICATION NUMBER: AU PP6061  
/ PRIOR FILING DATE: 1998-09-22  
/ PRIOR APPLICATION NUMBER: AU PP6062  
/ PRIOR FILING DATE: 1998-09-22  
/ PRIOR APPLICATION NUMBER: AU PP6063  
/ PRIOR FILING DATE: 1998-09-22  
/ PRIOR APPLICATION NUMBER: AU PP6064  
/ PRIOR APPLICATION NUMBER: AU PP6065

PRIOR FILING DATE: 1999-07-01  
PRIOR APPLICATION NUMBER: AU P01346  
PRIOR FILING DATE: 1999-07-01  
NUMBER OF SEQ ID NOS: 239  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 225  
LENGTH: 746  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: EHZ2 peptide  
OTHER INFORMATION: Fragment  
US-10-231-778-225

Query Match 99.8%; Score 3802; DB 4; Length 746;  
Best Local Similarity 99.9%; Pred. No. 2.2e-265;  
Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGQTGKSKRGKPVCKRKVKSEYMRLOLKRFRADDEVKSMFSSNRKOKLERTETILNOEM 60  
DB 1 MGQTGKSKRGKPVCKRKVKSEYMRLOLKRFRADDEVKSMFSSNRKOKLERTETILNOEM 60  
QY 61 KORRIOPVHILTSVSLRGTRSCVTSDDLPTQVIPLKTLNAVASVPIYMSWSPLOQNF 120  
DB 61 KORRIOPVHILTSVSLRGTRSCVTSDDLPTQVIPLKTLNAVASVPIYMSWSPLOQNF 120  
QY 121 WEDETVLHNPYMGDEVLDQGTPIEBELIKNYDGKVGHDRCGFINDBI FVELVNALQ 180  
DB 121 WEDETVLHNPYMGDEVLDQGTPIEBELIKNYDGKVGHDRCGFINDBI FVELVNALQ 180  
QY 181 YNDDDDDDGDDPBEREERKQKLEDRDKESRPPKPSDKIFRAISSMPDKGTAEEL 240  
DB 181 YNDDDDDDGDDPBEREERKQKLEDRDKESRPPKPSDKIFRAISSMPDKGTAEEL 240  
QY 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREOSLSFTLLFCRGCFTYDCFLAPPH 300  
DB 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREOSLSFTLLFCRGCFTYDCFLAPPH 300  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360  
DB 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360  
QY 361 NSRSTPTINVLKSDTSDRAGETGGENNDKEERKODETSSSEANRCOTPIKM 420  
DB 361 NSRSTPTINVLKSDTSDRAGETGGENNDKEERKODETSSSEANRCOTPIKM 420  
QY 421 KENIBEPENVEWSGAASMFVLIIGTYNFCALIAELIGTKRCQYERFVKESSIIADA 480  
DB 421 KENIBEPENVEWSGAASMFVLIIGTYNFCALIAELIGTKRCQYERFVKESSIIADA 480  
QY 481 PAEDVTPPRKKRKRRLWAHCRKIQLKDGSSNHNVNYOPCDHBRQPCDSSCPCTIAQ 540  
DB 481 PAEDVTPPRKKRKRRLWAHCRKIQLKDGSSNHNVNYOPCDHBRQPCDSSCPCTIAQ 540  
QY 541 NCEKRCQSSSECQNRPFCCRCCKAQCNTQCCRYLAVERCDPDLCTLCAAHMWSKINS 600  
DB 541 NCEKRCQSSSECQNRPFCCRCCKAQCNTQCCRYLAVERCDPDLCTLCAAHMWSKINS 600  
QY 601 CKNCSIORGSKGHLAPADVAGWGIPIQPOKNEFISYCGEITISQDEADRGVYDK 660  
DB 601 CKNCSIORGSKGHLAPADVAGWGIPIQPOKNEFISYCGEITISQDEADRGVYDK 660  
QY 661 YWCSFLFNALNDFVVDATRKANKIRFANHSVNPNCYAKV 699  
DB 661 YWCSFLFNALNDFVVDATRKANKIRFANHSVNPNCYAKV 699

RESULT 4  
US-10-295-027-710  
Sequence 710, Application US/10295027  
Publication No. US2003023250A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel

APPLICANT: Aziz, Natsaba  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezl, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Ros Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/10/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PAM.

NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 710  
LENGTH: 746  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-295-027-710

Query Match 99.8%; Score 3802; DB 4; Length 746;  
Best Local Similarity 99.9%; Pred. No. 2.2e-265;  
Matches 698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGQTGKSKRGKPVCKRKVKSEYMRLOLKRFRADDEVKSMFSSNRKOKLERTETILNOEM 60  
DB 1 MGQTGKSKRGKPVCKRKVKSEYMRLOLKRFRADDEVKSMFSSNRKOKLERTETILNOEM 60  
QY 61 KORRIOPVHILTSVSLRGTRSCVTSDDLPTQVIPLKTLNAVASVPIYMSWSPLOQNF 120  
DB 61 KORRIOPVHILTSVSLRGTRSCVTSDDLPTQVIPLKTLNAVASVPIYMSWSPLOQNF 120  
QY 121 WEDETVLHNPYMGDEVLDQGTPIEBELIKNYDGKVGHDRCGFINDBI FVELVNALQ 180  
DB 121 WEDETVLHNPYMGDEVLDQGTPIEBELIKNYDGKVGHDRCGFINDBI FVELVNALQ 180  
QY 181 YNDDDDDDGDDPBEREERKQKLEDRDKESRPPKPSDKIFRAISSMPDKGTAEEL 240  
DB 181 YNDDDDDDGDDPBEREERKQKLEDRDKESRPPKPSDKIFRAISSMPDKGTAEEL 240  
QY 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREOSLSFTLLFCRGCFTYDCFLAPPH 300  
DB 241 KEKYKELTEQOLPGALPPECTPNIDGNPAKSVQREOSLSFTLLFCRGCFTYDCFLAPPH 300  
QY 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360  
DB 301 ATPNTYKRNKTETALDNKPCGPOCYOHLGAKFPAALTAERIKTPPKRPGRRGRRLPN 360  
QY 361 NSRSTPTINVLKSDTSDRAGETGGENNDKEERKODETSSSEANRCOTPIKM 420  
DB 361 NSRSTPTINVLKSDTSDRAGETGGENNDKEERKODETSSSEANRCOTPIKM 420



Db	361	NSRSTPTINVLJESKOTDSDREAGTBTGJENNDDKEEKKODETSSSSXANSRCQPIIK	420
Qy	421	KPNLEPPBNVWESGABASMFRLIGTYYNPCALRLITCKROYREPRVWESSITABA	480
Db	421	KPNLEPPBNVWESGABASMFRLIGTYYNPCALRLITCKROYREPRVWESSITABA	480
Qy	481	PAEDVDTPPRKKQKGRHMAAHCKRIQLTKDGSNMHVYNYQPCHDRQPCDSSCPVIAQ	540
Db	481	PAEDVDTPPRKKQKGRHMAAHCKRIQLTKDGSNMHVYNYQPCHDRQPCDSSCPVIAQ	540
Qy	541	NPCBEPCCOSSSBCONRPFQCRCKAOCNTKOCPCYLAVRBCDPLCLTCGAADHMDSKANS	600
Db	541	NPCBEPCCOSSSBCONRPFQCRCKAOCNTKOCPCYLAVRBCDPLCLTCGAADHMDSKANS	600
Qy	601	CKNCSIQGSKKGLHLLAPSDVAGWGJFIPIQVQKNEFISEYCGEIIISQDEADRGKAVYDK	660
Db	601	CKNCSIQGSKKGLHLLAPSDVAGWGJFIPIQVQKNEFISEYCGEIIISQDEADRGKAVYDK	660
Qy	661	YMCSTFLPMLNNDPVYDATKGNKIRPANSVNPNCYAKV	699
Db	661	YMCSTFLPMLNNDPVYDATKGNKIRPANSVNPNCYAKV	699

```

: RESULT 5
: US-10-231-778-226
: Sequence 226, Application US/10231778
: Publication No. US20030126647A1
: GENERAL INFORMATION:
: APPLICANT: Bildeau, Pierre
: APPLICANT: Chaudhury, Abdul M.
: APPLICANT: Dennis, Elizabeth S.
: APPLICANT: Koltunow, Anna M.G.
: APPLICANT: Luo, Ming
: APPLICANT: Peacock, William J.
: TITLE OF INVENTION: Method for inducing seed development by down-regulating
: TITLE OF INVENTION: expression of the F1S2 gene
: FILE REFERENCE: 72-98A
: CURRENT APPLICATION NUMBER: US/10/231,778
: CURRENT FILING DATE: 2002-11-08
: PRIOR APPLICATION NUMBER: 09/398,237
: PRIOR FILING DATE: 1999-09-20
: PRIOR APPLICATION NUMBER: 60/101,184
: PRIOR FILING DATE: 1998-09-21
: PRIOR APPLICATION NUMBER: AU PP6061
: PRIOR FILING DATE: 1998-09-22
: PRIOR APPLICATION NUMBER: AU PP6062
: PRIOR FILING DATE: 1998-09-22
: PRIOR APPLICATION NUMBER: AU PP6063
: PRIOR FILING DATE: 1998-09-22
: PRIOR APPLICATION NUMBER: AU P01345
: PRIOR FILING DATE: 1999-07-01
: PRIOR APPLICATION NUMBER: AU P01346
: PRIOR FILING DATE: 1999-07-01
: NUMBER OF SEQ ID NOS: 239
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 226
: LENGTH: 746
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Ezhi peptide
: US-10-231-778-226

```

Query Match	98.5%	Score 3750	DB 4	Length 746
Best Local Similarity	98.1%	Pred. No. 1.3e-261		
Matches 686; Conservative	7	Mismatches 6	Indels 0	Gaps 0

[illegible]

Db	61	KQRRIQPHIMISVSLSLGTRECVTDLDPRPAVILPLKTLNMAVASPIRMYNSPLQONF	120
Qy	121	MWEDFTVLHNIIPYMGDEVLDQGTPIBELIKNYDGKYGHGREGFIINDELFVBLVNAALQ	180
Db	121	MWEDFTVLHNIIPYMGDEVLDQGTPIBELIKNYDGKYGHGREGFIINDELFVBLVNAALQ	180
Qy	181	YNDDDDDDDGGDDPEEREEKOKDLEHNRDCKSPRPYFSPDKLFEALISMFDPKTAEBEL	240
Db	181	YNDDDDDDDGGDDPEEREEKOKDLEHNRDCKSPRPYFADKLFEMISSMFDPKTAEBEL	240
Qy	241	KEKXYELTBQULPGALPPECTPNIIDGNPANSVQREBSLHSPHTLFCRCKXYCTGFLPHH	300
Db	241	KEKXYELTBQULPGALPPECTPNIIDGNPANSVQREBSLHSPHTLFCRCKXYCTGFLPHH	300
Qy	301	AAPNTYKRNKETALDNPCCPOCYOHLEAGKEPAALTLBARTKTPPKRGRRRLPLN	360
Db	301	AIPNTYKRNKITALDNPCCPOCYOHLBAGKEPAALTLBARTKTPPKRGRRRLPLN	360
Qy	361	NSSRPSTPTINVLBSKDTDSDBRAGTETGGENNDKEEBKKDETSSSEANSFCQPIPK	420
Db	361	NSSRPSTPTINVLBSKDTDSDBRAGTETGGENNDKEEBKKDETSSSEANSFCQPIPK	420
Qy	421	KPIPIPEBVEVMSGABASMFYLIGTYVDFCAIARLIGTKCQOYVFPVKSSITIAPA	480
Db	421	KPIPIPEBVEVMSGABASMFYLIGTYVDNCAIARLIGTKCQOYVFPVEMSSITIAV	480
Qy	481	PAEDVPTPRKKKKRGRLLMAHCRKIQLKODGSSNHVYNTOPCDHPROPCDSSCPVIAQ	540
Db	481	PTEDVPTPRKKKKRGRLLMAHCRKIQLKODGSSNHVYNTOPCDHPROPCDSSCPVIAQ	540
Qy	541	NPEKFCOCCSSCONRPFGCFCRKAQCMTKOCPCLAVRECPDPLCLCGAAMHMSKNYS	600
Db	541	NPEKFCOCCSSCONRPFCRCRKAQCMTKOCPCLAVRECPDPLCLCGAAMHMSKNYS	600
Qy	601	CKNCSIORGSKGHLILAPSDVAGWGITFKDPVOQKNEPISBYCBIISQDEADRRGKYDK	660
Db	601	CKNCSIORGSKGHLILAPSDVAGWGITFKDPVOQKNEPISBYCBIISQDEADRRGKYDK	660
Qy	661	YMSGFLFNANDPVUDATRKGNKILRFAHSHVNPVCYAKV	699
Db	661	YMSGFLFNANDPVUDATRKGNKILRFAHSHVNPVCYAKV	699

```

RESULT 6
US-10-104-047-3162
; Sequence 3162, Application US/10104047
; Publication NO. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3162
; LENGTH: 707
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-104-047-3162

```

Query Match	93.7%	Score 3569.5	DB 4	Length 707
Best Local Similarly	94.3%	Pred. NC. 1.2e-248		
Matches 659	0	Mismatches 1	Indels 39	Gaps 1

QY 1 MGCTGKKSEKGPVCKARKYKSEYMFRLRQKPRADBYKXSFSSNRQKILBERTILNQW 60

Db 1 MGCTGKKSEKGPVCKARKYKSEYMFRLRQKPRADBYKXSFSSNRQKILBERTILNQW 60

QY 61 KQRRIQPVHLITSVSLRGTRGCSVTSDLPFTQYIPLKTLNAVASVPIMWSNGLQNF 120



```
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 354
LENGTH: 517
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-354

Query Match      68.3%; Score 2602; DB 4; Length 517;
Best Local Similarity 100.0%; Pred. No. 5,2e-179;
Matches 470; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 MPPDGTABELKKEKLETEQOLPGALPPECPTNIDGNVAVSVOEGSLHSFTLLFCGRG 289
DB 1 MPPDGTABELKKEKLETEQOLPGALPPECPTNIDGNVAVSVOEGSLHSFTLLFCGRG 60
QY 290 FKYDCLFPHATPTTYRKKTETALDNKPCGPOCYOHLGAKFPAALTLARIKTPPKR 349
DB 61 FKYDCLFPHATPTTYRKKTETALDNKPCGPOCYOHLGAKFPAALTLARIKTPPKR 120
QY 350 PGGRRRGLPNNSSPSTPTTIVLESKOTSDREAGTGTGCHNDKSEBKDETSSESSE 409
DB 121 PGGRRRGLPNNSSPSTPTTIVLESKOTSDREAGTGTGCHNDKSEBKDETSSESSE 180
QY 410 ANSRQOTPIKMPNIEPPNENVMGASAMFVLLGTYVNDCAIARLIGTTCQVYEF 469
DB 181 ANSRQOTPIKMPNIEPPNENVMGASAMFVLLGTYVNDCAIARLIGTTCQVYEF 240
QY 470 RVKSSIIAPAPADVDTPPRKKKKRHLMAHCKRIQLKXGSSNNHYVYQPCDHPQP 529
DB 241 RVKSSIIAPAPADVDTPPRKKKKRHLMAHCKRIQLKXGSSNNHYVYQPCDHPQP 300
QY 530 CDSGCPVIAQNFCEKFCQSSSECONRFPGCRCAQACNTKQCPCTLAARECDPDLCTCG 589
DB 301 CDSGCPVIAQNFCEKFCQSSSECONRFPGCRCAQACNTKQCPCTLAARECDPDLCTCG 360
QY 590 AADHNDKSNVSGKNSIORGSKHLLASDVAGWGIPLKDPVQKNEFISECGELISOD 649
DB 361 AADHNDKSNVSGKNSIORGSKHLLASDVAGWGIPLKDPVQKNEFISECGELISOD 420
QY 650 EADRGKVDYDKYMGFLFYLANDFYVDATRKGNKIRFANHSVNPICYAV 699
DB 421 EADRGKVDYDKYMGFLFYLANDFYVDATRKGNKIRFANHSVNPICYAV 470

RESULT 9
US-10-115-482-56
Sequence 56, Application US/10115482
Publication No. US20030212257A1
GENERAL INFORMATION:
APPLICANT: Syntek, et al.
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
TITLE OF INVENTION: AND METHODS
TITLE OF INVENTION: OF USING THE SAME
FILE REFERENCE: 21404-322D
CURRENT APPLICATION NUMBER: US/10/115,482
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: 60/281,086
```

```
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281,136
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281,863
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/281,906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282,934
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283,512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/285,325
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: 60/285,890
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 60/286,068
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/286,292
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/287,213
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: 60/288,257
PRIOR FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: 60/291,134
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/282,020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/291,725
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/294,771
PRIOR FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/296,965
PRIOR FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: 60/299,128
PRIOR FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 149
SEQ ID NO 56
LENGTH: 723
TYPE: PRT
ORGANISM: Homo sapiens
US-10-115-482-56

Query Match      58.1%; Score 2212.5; DB 4; Length 723;
Best Local Similarity 59.7%; Pred. No. 1e-150;
Matches 425; Conservative 90; Mismatches 120; Indels 77; Gaps 11;

QY 15 WRKRVKSEYMLRQLKRRFRADRVKSMFSSNRKILERTETINQEMKORRLOPVHILTSV 74
DB 15 WRKRVKSEYMLRQLKRRFRADRVKSMFSSNRKILERTETINQEMKORRLOPVHILTSV 74
QY 75 SGLRGRECSVTSDL-DEPTGVILKTLNVAASVPIMYSNPLOQNFVDETYLANIPY 133
DB 75 SGLRGRECSVTSDL-DEPTGVILKTLNVAASVPIMYSNPLOQNFVDETYLANIPY 134
QY 134 MGDEYLDODGFIFIELIKNYGKYHGRE----CGFINDEL FVELVNALGOYND----- 185
DB 135 MGDEYKEDDEFIEBELINNYDGKVGEBEMTGSVLSDAVLELVDLNGLYSDEEBEGH 194
QY 186 -----DDDDGDGDPREERERKOKOLBDHDDKESRPPRKFPSPDKIFRALISMPEDKSTA 237
DB 195 NDTSDGKDDSKEDLPVTRKKRAHIBGNK--KSK--KQFPNDMIFALISMPENGV 250
QY 238 BELKRYKLETEQOLPGALPPECPTNIDGNVAVSVOEGSLHSFTLLFCRCFCYDCFLH 297
DB 251 DDMKERYRELTEMSPNLPQCTPTNIDGNVAVSVOEGSLHSFTLLFCRCFCYDCFLH 310
QY 298 PPHATPTTYRKKTETALDNKPCGPOCYOHLGAKFPAALTLARIKTPPKRPGRRGR 357
DB 311 PPHATPTTYRKKTETALDNKPCGPOCYOHLGAKFPAALTLARIKTPPKRPGRRGR 364
QY 358 LPNNSRSTPTIN-VLESKOTSDREAGTGTGCHNDKSEBKDETSSESSEANSRCOT 416
DB 365 HHIVSASCSNAGASAVATKGGSDSDYTG-----NDMASSSEANSRCOT 409
```

```

Qy 417 P|K|K|E|N|I|P|-----E|N|E|S|G|A|S|M|F|R|V|L|G|T|Y|D|N|F|C|A|I|A|L|I|G|K|T|R|O|V|Y 467
Db 410 P|K|Q|A|S|P|A|P|O|L|C|V|E|A|D|S|E|P|E|W|T|G|A|E|S|L|F|R|V|H|G|T|Y|F|N|N|F|C|I|A|L|I|G|K|T|C|Q|V|P 469
Qy 468 B|E|R|V|E|S|S|I|I|A|P|A|D|V|D|T|P|R|K|K|R|K|R|L|A|A|H|C|R|K|I|Q|L|K|D|G|S|N|H|V|N|Y|Q|P|C|D|H|R 527
Db 470 Q|A|V|E|S|S|I|I|A|P|A|D|V|D|T|P|R|K|K|R|K|R|L|A|A|H|C|R|K|I|Q|L|K|D|G|S|N|H|V|N|Y|Q|P|C|D|H|R 528
Qy 528 Q|P|C|D|S|C|P|V|I|A|N|F|C|E|K|T|C|Q|C|S|S|E|C|O|N|R|P|G|C|R|C|A|O|C|N|T|Q|C|P|C|T|A|V|R|E|C|D|P|L|C|T 587
Db 529 R|E|C|D|S|C|P|C|I|M|T|O|N|F|C|E|K|T|C|Q|C|N|P|D|-----L|R|E|C|D|P|L|C|T 564
Qy 588 C|S|A|H|H|W|D|S|K|N|V|S|C|R|K|C|S|I|O|R|G|S|K|R|L|L|A|P|S|D|V|A|G|M|G|I|F|I|D|P|V|O|K|N|E|I|S|E|Y|C|E|I|S 647
Db 565 C|A|S|E|R|H|D|C|K|V|V|S|C|R|K|C|S|I|O|R|G|L|K|R|L|L|A|P|S|D|V|A|G|M|G|I|F|I|K|S|V|O|K|N|E|I|S|E|Y|C|E|I|S 624
Qy 648 Q|D|E|A|D|R|G|K|V|Y|K|Y|K|Y|C|S|F|L|F|N|L|N|D|F|V|V|D|A|T|R|K|G|K|I|R|P|A|N|S|V|N|P|N|C|Y|A|K|V 699
Db 625 Q|D|E|A|D|R|G|K|V|Y|K|Y|K|Y|C|S|F|L|F|N|L|N|D|F|V|V|D|A|T|R|K|G|K|I|R|P|A|N|S|V|N|P|N|C|Y|A|K|V 676

RESULT 10
US-11-097-143-14073
; Sequence 14073, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14073
; LENGTH: 760
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-14073

Query Match 51.5%; Score 1959.5; DB 6; Length 760;
Best Local Similarity 52.2%; Pred No. 1.9e-132;
Matches 394; Conservative 84; Mismatches 170; Indels 107; Gaps 14;

```

```

Db 116 B|E|T|V|L|H|N|I|P|Y|W|G|D|E|V|L|D|K|G|F|I|E|L|I|K|N|Y|D|K|V|H|G|D|K|P|S|F|M|D|A|I|P|E|L|V|H|A|L|M|S|Y 175
Qy 179 -----G|Q|Y|N|D|-----D|D|D|D|G|D|P|-----B|E|R|E 198
Db 176 S|K|E|L|E|A|A|P|G|T|A|T|A|K|T|E|L|T|A|K|S|Q|G|E|D|G|V|D|V|A|D|E|S|P|M|K|L|E|R|T|D|S|K|G|L|T|E|V|E|R|K|E 235
Qy 199 K|O|K|O|E|D|R|U|D|-----K|E|S|R|P|R|K|P|S|D|K|I|F|E|A|I|S|W|P|D|K|T|A|B|E|L|K|E|K|Y|E|L|E|T|O|O|P 253
Db 236 T|E|P|L|E|T|D|A|D|V|K|P|V|E|V|K|D|L|P|P|P|A|I|I|F|O|A|I|N|A|N|P|D|K|T|A|O|L|E|K|Y|I|E|L|E|H|O|P 295
Qy 254 G|A|L|P|E|C|T|P|N|I|D|G|P|N|A|K|S|V|O|R|E|O|S|I|S|F|H|L|J|C|R|B|C|F|Y|D|C|E|L|A|P|P|-----H|A|T|E|N|T|Y|R|K|N|T 311
Db 296 -E|R|P|E|C|T|P|N|I|D|G|I|A|B|S|V|R|R|E|R|T|H|S|F|H|L|J|C|R|B|C|F|Y|D|C|E|L|A|P|P|-----H|A|T|E|N|T|Y|R|K|N|T 354
Qy 312 E|T|A|L|N|K|C|G|P|O|C|Y|H|L|G|A|N|G|F|A|A|L|T|A|R|I|K|T|P|P|K|R|P|G|R|R|R|G|L|P|N|N|S|R|P|S|T|P|T|N 371
Db 355 E|L|K|P|R|A|B|C|N|S|C|Y|N|L|D|G|M|E|K|L|A|A|D|-----K|T|P|----- 386
Qy 372 V|L|E|S|K|D|T|S|D|R|A|G|E|T|G|E|N|D|K|-----B|E|E|K|O|E|T|S|S|E|A|N|S|R|C|O|T|P|R|K|K|P|N|I 424
Db 387 -----I|D|S|C|N|E|A|S|E|D|S|N|S|O|S|N|O|P|N|H|E|N|S|K|D|N|G|L|V|N|S|A|V|A|E|I|N|S|I|N|G|A|N|N|I 440
Qy 425 B|E|P|E|N|E|W|S|G|A|E|S|M|F|R|V|L|G|T|Y|D|N|F|C|A|I|A|L|I|G|K|T|R|O|V|Y|E|R|V|E|S|S|I|I|A|P|A|D 484
Db 441 T|S|T|Q|C|V|-----W|T|G|A|D|Q|A|L|Y|R|V|L|H|V|Y|L|K|N|Y|C|A|I|A|H|N|L|T|K|R|O|V|Y|E|R|F|A|Q|E|D|A|F|S|E|D|R|Q 499
Qy 485 V|D|T|P|R|K|K|R|K|R|L|A|A|H|C|R|K|I|Q|L|K|D|G|S|N|H|V|N|Y|Q|P|C|D|H|R|O|P|C|D|S|C|P|V|I|A|N|F|C|E 544
Db 500 D|F|T|P|R|K|K|R|K|R|L|A|A|H|C|R|K|I|Q|L|K|D|G|S|N|H|V|N|Y|Q|P|C|D|H|R|O|P|C|D|S|C|P|V|I|A|N|F|C|E 559
Qy 545 K|P|C|Q|S|S|E|C|O|N|R|P|G|C|R|C|A|O|N|T|Q|C|P|C|Y|A|V|R|E|C|D|P|L|C|T|G|A|H|H|W|D|S|K|N|V|S|C|R|K|C|S| 604
Db 560 K|P|C|N|S|S|D|C|O|N|R|P|G|C|R|C|A|O|N|T|Q|C|P|C|Y|A|V|R|E|C|D|P|L|C|A|C|S|-----A|O|P|K|L|T|K|T|Y|C|N|V 618
Qy 605 S|I|O|R|G|S|K|R|L|L|A|P|S|D|V|A|G|M|G|I|F|I|D|P|V|O|K|N|E|I|S|E|Y|C|E|I|S|O|D|E|A|D|R|G|K|V|Y|K|Y|C|S 664
Db 619 C|V|O|R|G|H|H|L|A|P|S|D|V|A|G|M|G|I|F|L|K|E|G|A|Q|K|N|E|I|S|E|Y|C|E|I|S|O|D|E|A|D|R|G|K|V|Y|K|Y|C|S 678
Qy 665 F|L|F|N|L|N|D|F|V|V|D|A|T|R|K|G|K|I|R|P|A|N|S|V|N|P|N|C|Y|A|K|V 699
Db 679 F|L|F|N|L|N|D|F|V|V|D|A|T|R|K|G|K|I|R|P|A|N|S|V|N|P|N|C|Y|A|K|V 713

RESULT 11
US-10-231-778-224
; Sequence 224, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
; APPLICANT: Blodeau, Pierre
; APPLICANT: Chaudhury, Abdul M.
; APPLICANT: Dennis, Elizabeth S.
; APPLICANT: Koltunow, Anna M.G.
; APPLICANT: Luo, Ming
; APPLICANT: Peacock, William J.
; TITLE OF INVENTION: Method for inducing seed development by down-regulating
; TITLE OF INVENTION: expression of the P1s2 gene
; FILE REFERENCE: 72-98A
; CURRENT APPLICATION NUMBER: US/10/231,778
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: 09/398,237
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 60/101,184
; PRIOR FILING DATE: 1998-09-21
; PRIOR APPLICATION NUMBER: AU P6061
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU P6062
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU P6063
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU P61345
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: AU P61346
; PRIOR FILING DATE: 1999-07-01

```

; NUMBER OF SEQ ID NOS: 239  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 224  
 ; LENGTH: 760  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURES:  
 ; OTHER INFORMATION: Description of Artificial Sequence: E(2) peptide  
 ; OTHER INFORMATION: fragment  
 US-10-231-778-224

Query Match 51.3%; Score 1952.5; DB 4; Length 760;  
 Best Local Similarity 52.2%; Pred. No. 6.1e-132;  
 Matches 394; Conservative 84; Mismatches 170; Indels 107; Gaps 15;

```

QY      8 SEKPVCKRKVKSEYKRLRLKRRADVKSMSSNRKILERTL---NDEWQR 64
      3 STKVPPEKRRVKSEYKIRQKKRYKRADEIKKAWIRWDEHNHNVQDLYESKVMQKP 62
      65 IQPVHILTSVSLRGTSVTSDDLDPPT--QVPLKLTANAASVPIMYSNPLQONFMV 122
      63 YDPFHV-----DCVKRAEYTSYNGIPSGQKVPICVIMAVTPIPMYTAAPQONFMV 115
      123 EDETVLHNIIPYMGDEVLDQDGTPIBELIKNYDGVHGRBCGFINDEIFVELVNAL---- 178
      116 EDETVLHNIIPYMGDEVLDQDGTPIBELIKNYDGVHGRBCGFINDEIFVELVNALMSY 175
      179 -----GOYND---DDDDGDDEP--EREKQDLBD----- 205
      176 SKELBAAPSTSTAIKTEPLAKSKQGEDGVVDVADCSPMKLEKTSKGLTDVEKKE 235
      206 -----HRDDKESRPPRK-----PPSDKIFPAISMPEDKGTAELEKYELEEQLP 253
      226 TBEVETEDADVKAVEYVKDKLPFPAPITFOAISANPPDKGTAELEKYELEHQDP 295
      254 GALPPECTPNIDGNPAKSVQREQLSHSFHTLFCRCFCFYKDCFLHP--HATPNYTKRKT 311
      226 -ERPQECTPNIDGIAESVSRRRTWMSFHTLFCRCFCFYKDCFLHRLQGHAGPNLOKRRYP 354
      312 ETALDNKCGQCYOHLGAGKPAALTAERIKTTPKRGGRRRRLPNNSRSTPTIN 371
      355 ELKPAEPCSNSCYMLIDGMEKELAADS---KTPP----- 386
      372 VLESKDTSDREAGTETGENDK-----EEBEKDETSSSSEANSRCQPIIMKENI 424
      387 -----IDSCNEIASSEDSNDNSQPSNDFNHENSKONGLTVNSAAVAEINSIMAGMNI 440
      425 EPPENWMSGABASMFRLIGTYDNFCALIALIGTKTCROYEFRVYESSIIAPAPAD 484
      441 TSTQCV-WTGADQALYRLVLAHKYLLKNYCALAHNMULTKTCROYEFAQKEDAEFEDLRQ 499
      485 VTPPRKKKKRKRMAAHCRKIQLKKDSSNNVYVYQPCDHRORCDSSCPVIAONPCE 544
      500 DTPPRKKKKRKRMAAHCRKIQLKKDSSNNVYVYQPCDHRORCDSSCPVIAONPCE 559
      545 KFCQCSSECONFPGCRCAQCNTRQPCYLAVRBCDPLCLTCGAADHMSKNVSCNKC 604
      560 KFCNCSSDCONFPGCRCAQCNTRQPCYLAVRBCDPLCLTCGAADHMSKNVSCNKC 618
      605 STORGSKRLILAPSDVAGMGIPIYDPOVQNEPISYCEIISQDADRGRGVYDKMCS 664
      619 CVORGLHKLILAPSDIAGMGIPIYDPOVQNEPISYCEIISQDADRGRGVYDKMCS 678
      665 ELFNLANDPVNDATRKGNKIRPANSVNPNCYAKV 699
      679 FLFNLNDPVNDATRKGNKIRPANSVNPNCYAKV 713
  
```

RESULT 12  
 US-10-773-302-11  
 ; Sequence 11, Application US/10773302  
 ; Publication No. US20050089880A1  
 ; GENERAL INFORMATION:

; APPLICANT: Jenuwein, Thomas  
 ; APPLICANT: Laible, Gotz  
 ; APPLICANT: O'Carroll, Donal  
 ; APPLICANT: Eisenhaber, Frank  
 ; APPLICANT: Rea, Stephen  
 ; TITLE OF INVENTION: Chromatin-Regulator Genes  
 ; FILE REFERENCE: 0652.1670001  
 ; CURRENT APPLICATION NUMBER: US/10/773,302  
 ; PRIOR FILING DATE: 2004-02-09  
 ; PRIOR APPLICATION NUMBER: US/09/589,892  
 ; PRIOR FILING DATE: 2000-06-09  
 ; PRIOR APPLICATION NUMBER: US 08/945,988  
 ; PRIOR FILING DATE: 1997-11-10  
 ; PRIOR APPLICATION NUMBER: PCT/EP96/01818  
 ; PRIOR FILING DATE: 1996-05-02  
 ; PRIOR APPLICATION NUMBER: DE 195 16 776.7  
 ; PRIOR FILING DATE: 1995-05-10  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 11  
 ; LENGTH: 760  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 US-10-773-302-11

Query Match 50.8%; Score 1934.5; DB 5; Length 760;  
 Best Local Similarity 52.1%; Pred. No. 1.2e-130;  
 Matches 395; Conservative 85; Mismatches 165; Indels 113; Gaps 17;

```

QY      8 SEKPVCKRKVKSEYKRLRLKRRADVKSMSSNRKILERTL---NDEWQR 64
      3 STKVPPEKRRVKSEYKIRQKKRYKRADEIKKAWIRWDEHNHNVQDLYESKVMQKP 62
      65 IQPVHILTSVSLRGTSVTSDDLDPPT--QVPLKLTANAASVPIMYSNPLQONFMV 122
      63 YDPFHV-----DCVKRAEYTSYNGIPSGQKVPICVIMAVTPIPMYTAAPQONFMV 115
      123 EDETVLHNIIPYMGDEVLDQDGTPIBELIKNYDGVHGRBCGFINDEIFVELVNAL---- 178
      116 EDETVLHNIIPYMGDEVLDQDGTPIBELIKNYDGVHGRBCGFINDEIFVELVNALMSY 175
      179 -----GOYND---DDDDGDDEP--EREKQDLBD----- 205
      176 SKELBAAPSTSTAIKTEPLAKSKQGEDGVVDVADCSPMKLEKTSKGLTDVEKKE 235
      206 -----HRDDKESRPPRK-----PPSDKIFPAISMPEDKGTAELEKYELEEQLP 253
      226 TBEVETEDADVKAVEYVKDKLPFPAPITFOAISANPPDKGTAELEKYELEHQDP 295
      254 GALPPECTPNIDGNPAKSVQREQLSHSFHTLFCRCFCFYKDCFLHP--HATPNYTKRKT 311
      226 -ERPQECTPNIDGIAESVSRRRTWMSFHTLFCRCFCFYKDCFLHRLQGHAGPNLOKRRYP 354
      312 ETALDNKCGQCYOHLGAGKPAALTAERIKTTPKRGGRRRRLPNNSRSTPTIN 371
      355 ELKPAEPCSNSCYMLIDGMEKELAADS---KTPP----- 386
      372 VLESKDTSDREAGTETGENDK-----EEBEKDETSSSSEANSRCQPIIMKENI 424
      387 -----IDSCNEIASSEDSNDNSQPSNDFNHENSKONGLTVNSAAVAEINSIMAGMNI 440
      425 EPPENWMSGABASMFRLIGTYDNFCALIALIGTKTCROYEFRVYESSIIAPAPAD 484
      441 TSTQCV-WTGADQALYRLVLAHKYLLKNYCALAHNMULTKTCROYEFAQKED--AESEFD 496
      485 VD---TTPPRKKKKRKRMAAHCRKIQLKKDSSNNVYVYQPCDHRORCDSSCPVIAON 541
      497 LRODTPPRKKKKRKRMAAHCRKIQLKKDSSNNVYVYQPCDHRORCDSSCPVIAON 556
      542 FCEKFCQCSSECONFPGCRCAQCNTRQPCYLAVRBCDPLCLTCGAADHMSKNVSC 601
      557 FCEKFCNCSSDCONFPGCRCAQCNTRQPCYLAVRBCDPLCLTCGAADHMSKNVSC 615
  
```





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: August 11, 2006, 22:45:55 ; Search time 21.7682 Seconds  
(without alignments)  
2161.464 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGQTKSKSGKGVCMRKRVK.....KGNKIRFANHSVNPNCYAKV 699

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 239914 seqs, 67312017 residues

Total number of hits satisfying chosen parameters: 239914

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA New:\*  
1: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US09\_NEM\_PUB pep.\*  
2: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US06\_NEM\_PUB pep.\*  
3: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US07\_NEM\_PUB pep.\*  
4: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US08\_NEM\_PUB pep.\*  
5: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US10\_NEM\_PUB pep.\*  
6: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US11\_NEM\_PUB pep.\*  
7: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US11\_NEM\_PUB pep.\*  
8: /EMC\_Celerra\_SIDS3/Ptodata/2/pubpaa/US06\_NEM\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3787.5	99.5	751	US-10-539-630-1	Sequence 1, Appl
2	716	18.8	777	US-10-953-349-35469	Sequence 35469, A
3	716	18.8	895	US-10-953-349-35468	Sequence 35468, A
4	716	18.8	913	US-10-953-349-35467	Sequence 35467, A
5	697.5	18.3	894	US-10-953-349-37881	Sequence 37881, A
6	697.5	18.3	903	US-10-953-349-37880	Sequence 37880, A
7	694.5	18.2	874	US-10-953-349-37882	Sequence 37882, A
8	658	17.3	895	US-11-330-403-9227	Sequence 9227, A
9	232	6.1	511	US-10-953-349-35696	Sequence 35696, A
10	232	6.1	513	US-10-953-349-35695	Sequence 35695, A
11	232	6.1	551	US-10-953-349-35694	Sequence 35694, A
12	225.5	5.9	916	US-10-539-228-355	Sequence 355, A
13	218	5.7	497	US-11-330-403-16696	Sequence 16696, A
14	218	5.7	980	US-11-330-403-2871	Sequence 2871, A
15	210.5	5.5	870	US-10-539-228-358	Sequence 358, A
16	207	5.4	612	US-10-449-902-47628	Sequence 47628, A
17	206	5.4	425	US-11-330-403-11703	Sequence 11703, A
18	204	5.4	412	US-10-511-937-2580	Sequence 2580, A
19	203.5	5.3	954	US-11-330-403-11849	Sequence 11849, A
20	203.5	5.3	369	US-11-330-403-9402	Sequence 9402, A
21	203.5	5.3	786	US-11-330-403-1643	Sequence 1643, A
22	203	5.3	1212	US-10-449-902-46235	Sequence 46235, A
23	201.5	5.3	1051	US-11-330-403-8440	Sequence 8440, A
24	198.5	5.2	397	US-10-953-349-21996	Sequence 21996, A
25	198.5	5.2	407	US-11-330-403-6394	Sequence 6394, A

26	196	5.1	737	US-10-449-902-43109	Sequence 43109, A
27	195.5	5.1	352	US-11-330-403-14838	Sequence 14838, A
28	195.5	5.1	854	US-11-330-403-4780	Sequence 4780, A
29	193	5.1	354	US-10-953-349-21998	Sequence 21998, A
30	193	5.1	379	US-10-953-349-21997	Sequence 21997, A
31	192.5	5.1	585	US-11-330-403-5658	Sequence 5658, A
32	192.5	5.1	758	US-10-449-902-56544	Sequence 56544, A
33	192	5.0	344	US-11-330-403-2509	Sequence 2509, A
34	192	5.0	363	US-11-330-403-13204	Sequence 13204, A
35	191.5	5.0	684	US-11-330-403-6172	Sequence 6172, A
36	189.5	5.0	339	US-11-330-403-15444	Sequence 15444, A
37	189.5	5.0	343	US-11-330-403-13539	Sequence 13539, A
38	189	5.0	362	US-11-330-403-6904	Sequence 6904, A
39	188.5	5.0	337	US-11-330-403-10441	Sequence 10441, A
40	188	4.9	437	US-11-330-403-14362	Sequence 14362, A
41	188	4.9	870	US-11-330-403-3355	Sequence 3355, A
42	184	4.8	340	US-10-449-902-31539	Sequence 31539, A
43	184	4.8	477	US-11-330-403-14719	Sequence 14719, A
44	179.5	4.7	334	US-11-330-403-14432	Sequence 14432, A
45	179	4.7	340	US-11-330-403-5229	Sequence 5229, A

#### ALIGNMENTS

RESULT 1  
US-10-539-630-1  
; Sequence 1, Application US/10539630  
; Publication No. US20060104981A1  
; GENERAL INFORMATION:  
; APPLICANT: Takeda Chemical Industries, Ltd.  
; TITLE OF INVENTION: Preventing and treating agent for cancer  
; FILE REFERENCE: 3130WOOP  
; CURRENT APPLICATION NUMBER: US/10/539, 630  
; PRIOR FILING DATE: 2005-06-17  
; PRIOR APPLICATION NUMBER: JP2002-373144  
; PRIOR FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 14  
; SEQ ID NO 1  
; LENGTH: 751  
; TYPE: PRT  
; ORGANISM: Human  
; US-10-539-630-1

Query Match 99.5%; Score 3787.5; DB 6; Length 751;  
Best Local Similarity 99.1%; Pred. No. 3.5e-225;  
Matches 698; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY	1	MGQTKSKSGKGVCMRKRVKSYMLRQLKRRRADVYKSMSSNRKILERTILINDBW	60
DB	1	MGQTKSKSGKGVCMRKRVKSYMLRQLKRRRADVYKSMSSNRKILERTILINDBW	60
QY	61	KORRIQPVHILTSVSLNGTRECSTYSDLPFTQVILPLKTLNAVASVPIYMSWSPLOQNF	120
DB	61	KORRIQPVHILTSVSLNGTRECSTYSDLPFTQVILPLKTLNAVASVPIYMSWSPLOQNF	120
QY	121	MYEDTVLNIPIYMGDEVLDQGTIFIBLILNXYDKVHGDRGCFINDEIFVELYNALGQ	180
DB	121	MYEDTVLNIPIYMGDEVLDQGTIFIBLILNXYDKVHGDRGCFINDEIFVELYNALGQ	180
QY	181	YNDDDDDDDGDDPBERERKQDLRHRDCKSRPRKPPSDKIFALISSMFPDKGTABEL	240
DB	181	YNDDDDDDDGDDPBERERKQDLRHRDCKSRPRKPPSDKIFALISSMFPDKGTABEL	240
QY	241	KEKYKELTEOQLPGALPECTPNIDGNPAKSVORQSLHSFTTLFCRCFCYDCFLH---	297
DB	241	KEKYKELTEOQLPGALPECTPNIDGNPAKSVORQSLHSFTTLFCRCFCYDCFLH---	297
QY	301	NYSFATNTYKRNKTETALDNKPCGPOCYOHLBSAKAPPAALTAERTITPKRGCHRR	360
DB	301	NYSFATNTYKRNKTETALDNKPCGPOCYOHLBSAKAPPAALTAERTITPKRGCHRR	360
QY	356	GRLPNNSSRPSTPTINVLKSDTSDRKAAGTETGENNDDKEBEKCDFTSSSSSANSRQ	415



```
Db 361 |GRLPNNSGRSPSTPTINIVLESKDTSDBRAGTETGGENNDKEEKKEDTSSSSSEANSRCQ| 420
Qy 416 |TPIKKKPIBPENENWGSABSMFRVLIGTYNDFCALILGKTCKROYEPFVKKSS| 475
Db 421 |TPIKKKPIBPENENWGSABSMFRVLIGTYNDFCALILGKTCKROYEPFVKKSS| 480
Qy 476 |IAPAPADVDTPPRKKRKRHLMAAHCRIQLKKDGSNNHYNYQPCDHPRQPCDSSCP| 535
Db 481 |IAPAPADVDTPPRKKRKRHLMAAHCRIQLKKDGSNNHYNYQPCDHPRQPCDSSCP| 540
Qy 536 |CVIAONFCEKFCQCSSECONRPPGRCACQCTKCCPCYLAVERCDPDLCTCGAADHD| 595
Db 541 |CVIAONFCEKFCQCSSECONRPPGRCACQCTKCCPCYLAVERCDPDLCTCGAADHD| 600
Qy 596 |SKVNSCKRCSIORGSKKHLAPSDVAGMGIFIKDPVOXNPFISEYCGEIIISQDEADRG| 655
Db 601 |SKVNSCKRCSIORGSKKHLAPSDVAGMGIFIKDPVOXNPFISEYCGEIIISQDEADRG| 660
Qy 656 |KVYDKMCSFLFNLNDFVVDATRKGNKIRFANHSVNPVCYAKV| 699
Db 661 |KVYDKMCSFLFNLNDFVVDATRKGNKIRFANHSVNPVCYAKV| 704

RESULT 2
US-10-953-349-35469
; Sequence 35469, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35469
; LENGTH: 777
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35469

Query Match 18.8%; Score 716; DB 6; Length 777;
Best Local Similarity 28.2%; Pred. No. 2e-36;
Matches 220; Conservative 95; Mismatches 204; Indels 260; Gaps 35;

Qy 95 |VIPKLTINAASVPIIMYSWSPLOQN-FWVDETVL--HNIPY--MGDEVL---DQGTFTI| 146
Db 24 |VVRPVKPLVPIVERIPPTTWIFLDKNQRMADQSVGRRRIYYDPVGNBALICSDSD---| 79
Qy 147 |BELIKNYDGK--VHGD-----RBCGFINDIEFVELVNALGQYNDDDDDDDGDDPER| 195
Db 80 |EIEPPEBKKHPTFGEEOQLMRAWQEHG-LNREV---VAVLCQFID-----STPSE| 127
Qy 196 |REBKQKLEDRHDDKESRPPKPPSDKIFEAISMPDPKGTABELKKEYKELTEQQLPGA| 255
Db 128 |IEERSEVLFR-KNEKNSG-----SSDKI-----ERQL---| 153
Qy 256 |LPPECTPVIDGNPAKSVORBSLHSPFTLFCRRCFKYDCFLH-----PFAHPNTYK| 307
Db 154 |SLD-----KTM-----DAVLDSFQNLFCRCLVDFCDRLHGSQNLVFPTEKQPYSF-| 199
Qy 308 |RKNTETALDNKPCGPQCYQHLEG-----| 330
Db 200 |EPDENKKPCGRCQCTLMRGGFORIHVGLSGCATYNNESGTVSHKVDVSIIMSESD| 255
Qy 331 |AKEPAAALTABRIKTPP-----| 349
Db 256 |SNREKGNIRSMTLVGTSGSKTISSVASBESTTPPSADTSETENASSDMPPSLRKYKISK| 315
Qy 350 |PGRRRGGLPNNSSRPSPTPTI---NVL-----ESKQTDSDREAGTGTGGERNDKEEKE| 399
```

```
Db 316 |RGPYRERSPGKROKVTFTSDISFASNILNKLSIEIRDY---RLSEBPGG----DKLQ| 367
Qy 400 |KDETSSSSEANSRC-OTPIKKNPIEPEVE-----WSGAEMAFR| 441
Db 368 |ILDBSTKTSKSDICGSPITTEM---GIEBKVYSTGNFLEHTLSCMSALERDLYL| 423
Qy 442 |VLIGTYDNFCALIR-LIGTKROYEFVKKSSIIAPAP-----| 481
Db 424 |KGIEIFGKNSCLIRARNLISGKTCMEVANNNGAAMAKRPLINKSISGDPASTEODYM| 483
Qy 482 |AEDVTPPRKKRKR-----LMAHCRIQLKKDGSNNHY-NYOPCDHPRQPCDSSC| 534
Db 484 |EODVNAATRIYRRGRNRKIKTKYSAGHPTVRKRIEDGKQWYQYNFC-VQDMCGKDC| 542
Qy 535 |PCVIAONFCEKFCQCSSECONRPPGRCRC-KAQCNTKCCPCYLAVERCDPDLCTCGAADH| 593
Db 543 |PCVENGTCEKFCQCSSECONRPPGRCRCRC-KAQCNTKCCPCYLAVERCDPDLCTCGAADH| 597
Qy 594 |WDS-----KNVSCNCSIORGSKKHLAPSDVAGMGIFIKDPVOXNPFISE| 640
Db 598 |WVSCGDSIAGBPARGDGYCCANNKLLKQOQRIILGRSDVAGMGAFIKXNVNKNQDYLG| 657
Qy 641 |YCGRIISQDRADERGKYVDKTMCSFLFNLNDFVVDATRKGNKIRFANHSVNPVCYAKV| 699
Db 658 |YTGELISHKEDKRGKTYDRANSSFLPDLNDQVLDAYRGDKKLFANHSNPVCYAKV| 716

RESULT 3
US-10-953-349-35468
; Sequence 35468, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35468
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35468

Query Match 18.8%; Score 716; DB 6; Length 895;
Best Local Similarity 28.2%; Pred. No. 2.4e-36;
Matches 220; Conservative 95; Mismatches 204; Indels 260; Gaps 35;

Qy 95 |VIPKLTINAASVPIIMYSWSPLOQN-FWVDETVL--HNIPY--MGDEVL---DQGTFTI| 146
Db 142 |VVRPVKPLVPIVERIPPTTWIFLDKNQRMADQSVGRRRIYYDPVGNBALICSDSD---| 197
Qy 147 |BELIKNYDGK--VHGD-----RBCGFINDIEFVELVNALGQYNDDDDDDDGDDPER| 195
Db 198 |EIEPPEBKKHPTFGEEOQLMRAWQEHG-LNREV---VAVLCQFID-----STPSE| 245
Qy 196 |REBKQKLEDRHDDKESRPPKPPSDKIFEAISMPDPKGTABELKKEYKELTEQQLPGA| 255
Db 246 |IEERSEVLFR-KNEKNSG-----SSDKI-----ERQL---| 271
Qy 256 |LPPECTPVIDGNPAKSVORBSLHSPFTLFCRRCFKYDCFLH-----PFAHPNTYK| 307
Db 272 |SLD-----KTM-----DAVLDSFQNLFCRCLVDFCDRLHGSQNLVFPTEKQPYSF-| 317
Qy 308 |RKNTETALDNKPCGPQCYQHLEG-----| 330
Db 318 |EPDENKKPCGRCQCTLMRGGFORIHVGLSGCATYNNESGTVSHKVDVSIIMSESD| 373
Qy 331 |AKEPAAALTABRIKTPP-----| 349
Db 374 |SNREKGNIRSMTLVGTSGSKTISSVASBESTTPPSADTSETENASSDMPPSLRKYKISK| 433
```

QY 350 PGRRRGRLLPNNSSRPSTPTI---NVL-----ESKDTDSREAGTETGGENNDKEE 399  
DB 434 RGRYRERSPGKOKVFTSDISPASNIINKLSIPEIRDT---RLSERREG-----DKLQ 485  
QY 400 KODETSSSSSEANSRC--QPIKMKPNIBPENVE-----WGAEMAMR 441  
DB 486 IJDESTKTSKXKOCESPTITTEM---GIESKVSSTKNFLHTLSCWALERDYL 541  
QY 442 VLIGTYDNFCAIAR--LIGTKCYOYEFYKESIIAPAP-----481  
DB 542 KGIKIFGKNSCLIAINLISGMTKMEVANNYNNNGAMAKRPLINKSISGDPATEQDYM 601  
QY 482 AEDVDTPPKRRKRRH-----LMAHCKIOLKODGSSNNHY--NYOPCDHPQPCDSSC 534  
DB 602 EODMVARRTIYRRRGRNRKLTWTYSAGHPYTRKRIIDGKOWTYOYNPC--VCOQMGKDC 660  
QY 535 PCVIAONPCERKQCSCSBCQNRFPQGRG--KAOCNTKQCPCYLAVERCDPDLCLTGAADH 593  
DB 661 PCVENGTCCERKCGSKSCKNFRGCHCAKSQCRSRQCPFAASRECDPDVCRCN-----715  
QY 594 WDS-----KNVSCKNCSIQRGSKKHLILAPSDVAGWGIPIFDPOKNEFISE 640  
DB 716 WVSCGDSIGSEPPARGDGYCCGMKLLKQOORILLGRSDVAGWGAFIKIPVAKNDYLGE 775  
QY 641 YGCEIISODEADRGKVDKMYCSPLFINLNDPVNDATRKNIKIRPANSVNPNCYAKV 699  
DB 776 YTGELISHKADRGKRTYDRANSSFLFDLNDQYVLDAYRKGDKLKFXANSSNPNCYAKV 834

## RESULT 4

US-10-953-349-35467  
; Sequence 35467, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THEREBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 35467  
; LENGTH: 913  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35467

Query Match 18.8%; Score 716; DB 6; Length 913;  
Best Local Similarity 28.2%; Pred. No. 2,4e-36;  
Matches 220; Conservative 95; Mismatches 204; Indels 260; Gaps 35;  
QY 95 VPLKLTANVAVSPIMYSMPLOON--FMVDETVL--HNIPY--MGDEVY---DODGTPL 146  
DB 160 VVAFVKLPLVERIPPTTWIFLDKQRMADDSVGRRIIYDVGNEALICSDS-----215  
QY 147 BELIKNYDGK---VHGD-----RCCGFINDIIVEILVNALGQYNDDDDDGDDPBR 195  
DB 216 BELPPEBEKHFPTBEGDQIWRATQEHG--LNREV---VNVLCQFID-----STPSS 263  
QY 196 REKQKQDLBDHRDDKSRPPKRPSPDKIFEAISMPDPDKTABELAKTKALTEQOLPQA 255  
DB 264 IBERSEVLPE--KNEKNSG-----SSDKI-----ERQI-----289  
QY 256 LPECTPNTIDGPAKSVOREQSLSFHTLPCRCAPKYDCFIH-----PFAATYNTYK 307  
DB 290 -----SLD---KTM---DAVDSFDNLPFCRCILVDFCRILHCSQNLVPTFKQYSTR 335  
QY 308 RKNETALDNKPCGPOCYOHLG-----330  
DB 336 -----EPDENKCPGRCQCYLWRGGFOBIHDVGLSGCATYNNMESGTVSHKVDVIMSSED 391

QY 331 -----AKFAALTAERIKTP-----XR 349  
DB 392 SNREKGNIRSMTLVGTSSGKLISSVASESTTTPPSADTSETNASSDMPSSLRKTKISK 451  
QY 350 PGRRRGRLLPNNSSRPSTPTI---NVL-----ESKDTDSREAGTETGGENNDKEE 399  
DB 452 RGRYRERSPGKOKVFTSDISPASNIINKLSIPEIRDT---RLSERREG-----DKLQ 503  
QY 400 KODETSSSSSEANSRC--QPIKMKPNIBPENVE-----WGAEMAMR 441  
DB 504 IJDESTKTSKXKOCESPTITTEM---GIESKVSSTKNFLHTLSCWALERDYL 559  
QY 442 VLIGTYDNFCAIAR--LIGTKCYOYEFYKESIIAPAP-----481  
DB 560 KGIKIFGKNSCLIAINLISGMTKMEVANNYNNNGAMAKRPLINKSISGDPATEQDYM 619  
QY 482 AEDVDTPPKRRKRRH-----LMAHCKIOLKODGSSNNHY--NYOPCDHPQPCDSSC 534  
DB 620 EODMVARRTIYRRRGRNRKLTWTYSAGHPYTRKRIIDGKOWTYOYNPC--VCOQMGKDC 678  
QY 535 PCVIAONPCERKQCSCSBCQNRFPQGRG--KAOCNTKQCPCYLAVERCDPDLCLTGAADH 593  
DB 679 PCVENGTCCERKCGSKSCKNFRGCHCAKSQCRSRQCPFAASRECDPDVCRCN-----733  
QY 594 WDS-----KNVSCKNCSIQRGSKKHLILAPSDVAGWGIPIFDPOKNEFISE 640  
DB 734 WVSCGDSIGSEPPARGDGYCCGMKLLKQOORILLGRSDVAGWGAFIKIPVAKNDYLGE 793  
QY 641 YGCEIISODEADRGKVDKMYCSPLFINLNDPVNDATRKNIKIRPANSVNPNCYAKV 699  
DB 794 YTGELISHKADRGKRTYDRANSSFLFDLNDQYVLDAYRKGDKLKFXANSSNPNCYAKV 852

## RESULT 5

US-10-953-349-37881  
; Sequence 37881, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THEREBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 37881  
; LENGTH: 894  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-37881

Query Match 18.3%; Score 697.5; DB 6; Length 894;  
Best Local Similarity 25.6%; Pred. No. 3.3e-35;  
Matches 240; Conservative 103; Mismatches 240; Indels 353; Gaps 42;  
QY 3 QNGKSKSGK-----PVCMRKRVKSEMRRLKQLKFRARADVKSMSPSNOKLIE 51  
DB 12 QRSKSDQMGKDAASAASVPI-----HANLTOL-----IRQVGSGLAYIKR 54  
QY 52 RTEIILNDEWKOR---IOPVHILTSVSLRG-----RCSVTSDL-----89  
DB 55 KLEV--NRKTLQHSGLSPVAAAABVAS--RGTDGNAISQRAAEKQCG--SLANGIGER 110  
QY 90 -----DPTQVILPLKLTANA-----SVPIYMSMPLOON--FMVDETV 126  
DB 111 DVSVQBEMLATGIALSSGATAGRTIVRFVKLPLVERKIPPTTWIFLDKQRMADDS 170  
QY 127 VL--HNIPY--MGDEVY---DODGTPIE-----LIKRYDKV-----HDDRGGFTIN 167  
DB 171 VVGRRIIYDVGNEALICSDSDEBIPPEBEKHFPTGEBHLIWRATQDHG-----LN 224  
QY 168 DEIFVELVNALGQYNDDDDDGDDPBERERKQDLEHRRDDKSRPPKRPSS-----DKI 223

Db 225 QEV-----VNVLCOF-----IGATPSRIERSVLFPE-KNEKHSGSSDKIESRLSLDKT 272  
QY 224 FEAISMPDDKGTABELKEKTELTEQOLPGALPPECPTNIDGPAKSVQREQSLSHPT 283  
Db 273 MDAY-----LDSFDN 282  
QY 284 LFCRCFCYDCFLH-----PFIATPNTYKAKTETALDNKPCGPCY-----Q 326  
Db 283 LFCRCFLVDFCRHGCSONLVFPCEKQYSPDPBNK-----KPCGHLCTYLAPQ 332  
QY 327 HLBGAKE----- 333  
Db 333 WREGKEMHDDGLAGATYTWESGTAQORVDVNVNMYESEDNRQKNI RSMTLVGTSGSK 392  
QY 334 PAALITARITPP-----KPCGRRRGLPNNSRPSPT 369  
Db 393 IISVSABESTTTPADISETENVSSDLPPSSLRKHKISKQGPYRHSFGKQKQVFTSD 452  
QY 370 I-----NVLESKDTSDRBAGETEGENNDEBEKEDTSSSSSEANSRC-QTPI 418  
Db 453 ISFEGINMKLSIPIRDT--KLESRESG-----DKRLIDESTTKTSKMDGSESPA 504  
QY 419 KKKENIEBPPEN-----VE-----WSGABASMFRLIGTYTDFCAIAR--LIGTTC 463  
Db 505 TTMEVVGQSNKVYSTKNFLBSTLSCWSALBRDYLKGIIFGKNSCLIAHNLISGLTIC 564  
QY 464 ROYEFKYESIITAPAP-----ADVDTPPKKKRKR-----LM 499  
Db 565 IEVANYMYNNGAMAKRPLANKSISGDAFNEBQDMEQMAARTIYRRGNRKLKXTW 624  
QY 500 --AAHCRKIQLKDDSSNHVNYOPCDHPROPDCSVCVIAQNFCEKFCOCSSECONRF 557  
Db 625 KSAQH-PIYRKRTDGGKCTQYSPC-ACQMGCGDCPCADKGTCEKXCGSKSKXKF 682  
QY 558 PCGRG-KAQCNTKQCPCTYLAVERCDPDLCTCGAADHWS-----KNVSCN 603  
Db 663 RCHCAKSGQCSRQCPFAASRECDPDVCRNC-----WVSCGDGSLGEPPLARGDYOCGN 737  
QY 604 CSIQGSKKHLLAASDVAGMGIFIKDPVQKNEFISEYCGELISODBARBGKYDKWMC 663  
Db 738 MTLKLOOQORILLGSSDVAGWGAFLKNPVKNNDYIGETTGELISHKEDKRGKLTIDRANS 797  
QY 664 SFLFNLDNFVVDATRKGNKIRIFANHSVNPNCYAKV 699  
Db 798 SFLFDLNDQYVLDAYRKDKLKFANHSSNPNCYAKV 833

## RESULT 6

US-10-953-349-37880  
; Sequence 37880, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 37880  
; LENGTH: 903  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-37880

Query Match 18.3%; Score 697.5; DB 6; Length 903;  
Best Local Similarity 25.6%; Pred. No. 3.3e-35;  
Matches 240; Conservative 103; Mismatches 240; Indels 353; Gaps 42;  
QY 3 QGKSKSEK-----PVCWRKRVSEYMLRLQKRRFRADAVKMSFSSNRKQIIE 51

Db 21 QRSKRSQCGMKAAAAAIVPI-----HANTPL-----IROVSGRLAYIKE 63  
QY 52 RTEILINQEMQOR--IQPVHILTSVSSLGT-----RECSVTSDL----- 89  
Db 64 KLBV-NKRTIQRHSCSLFDVAAAARVAS-KGTQDGNALSORAARQCG--SDLANGIGER 119  
QY 90 -----DPEQVIPLKTLNAAV-----SVPIMVMSPLQON-FMVEDET 126  
Db 120 DVSVQEBENLATYTLALSSGATQRTIVRFVKPLVEKIPYTTWLFIDKQGMADQDS 179  
QY 127 VL--HNIPY--MGDEVL--DODGFTIE-----LIKNYDGKV-----HGDRCGRTN 167  
Db 180 VVGRRIIYDTVGNHALICSDSDEIIPREBEKHFPGEDHLMWRATQDHG-----LN 233  
QY 168 DEIFVELVNALGOYNDDDDDGDDPEREKKQLEDHRDDKESRPPKPS--DKI 223  
Db 234 QEV-----VNVLCOF-----IGATPSRIERSVLFPE-KNEKHSGSSDKIESRLSLDKT 281  
QY 224 FEAISMPDDKGTABELKEKTELTEQOLPGALPPECPTNIDGPAKSVQREQSLSHPT 283  
Db 282 MDAY-----LDSFDN 291  
QY 284 LFCRCFCYDCFLH-----PFIATPNTYKAKTETALDNKPCGPCY-----Q 326  
Db 292 LFCRCFLVDFCRHGCSONLVFPCEKQYSPDPBNK-----KPCGHLCTYLAPQ 341  
QY 327 HLBGAKE----- 333  
Db 342 WREGKEMHDDGLAGATYTWESGTAQORVDVNVNMYESEDNRQKNI RSMTLVGTSGSK 401  
QY 334 PAALITARITPP-----KPCGRRRGLPNNSRPSPT 369  
Db 402 IISVSABESTTTPADISETENVSSDLPPSSLRKHKISKQGPYRHSFGKQKQVFTSD 461  
QY 370 I-----NVLESKDTSDRBAGETEGENNDEBEKEDTSSSSSEANSRC-QTPI 418  
Db 462 ISFEGINMKLSIPIRDT--KLESRESG-----DKRLIDESTTKTSKMDGSESPA 513  
QY 419 KKKENIEBPPEN-----VE-----WSGABASMFRLIGTYTDFCAIAR--LIGTTC 463  
Db 514 TTMEVVGQSNKVYSTKNFLBSTLSCWSALBRDYLKGIIFGKNSCLIAHNLISGLTIC 573  
QY 464 ROYEFKYESIITAPAP-----ADVDTPPKKKRKR-----LM 499  
Db 574 IEVANYMYNNGAMAKRPLANKSISGDAFNEBQDMEQMAARTIYRRGNRKLKXTW 633  
QY 500 --AAHCRKIQLKDDSSNHVNYOPCDHPROPDCSVCVIAQNFCEKFCOCSSECONRF 557  
Db 634 KSAQH-PIYRKRTDGGKCTQYSPC-ACQMGCGDCPCADKGTCEKXCGSKSKXKF 691  
QY 558 PCGRG-KAQCNTKQCPCTYLAVERCDPDLCTCGAADHWS-----KNVSCN 603  
Db 692 RCHCAKSGQCSRQCPFAASRECDPDVCRNC-----WVSCGDGSLGEPPLARGDYOCGN 746  
QY 604 CSIQGSKKHLLAASDVAGMGIFIKDPVQKNEFISEYCGELISODBARBGKYDKWMC 663  
Db 747 MTLKLOOQORILLGSSDVAGWGAFLKNPVKNNDYIGETTGELISHKEDKRGKLTIDRANS 806  
QY 664 SFLFNLDNFVVDATRKGNKIRIFANHSVNPNCYAKV 699  
Db 807 SFLFDLNDQYVLDAYRKDKLKFANHSSNPNCYAKV 842

## RESULT 7

US-10-953-349-37882  
; Sequence 37882, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDE  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349

CURRENT FILING DATE: 2004-09-30  
 NUMBER OF SEQ ID NOS: 40252  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 37882  
 LENGTH: 874  
 TYPE: PRT  
 ORGANISM: Zea mays subsp. mays  
 US-10-953-349-37882

Query Match 18.2%; Score 694.5; DB 6; Length 874;  
 Best Local Similarity 26.1%; Pred. No. 4.9e-35;  
 Matches 232; Conservative 98; Mismatches 235; Indels 325; Gaps 39;

38 VASMSNSNKKLIEREILNCEKOR---IQPVHLTSVSSLRG-----RE 82  
 21 INOVSGRLAYIKKLEV-NRKTLOHSGSLPDVAAAASVAS-RGTDGNALSGRAESQ  
 83 CSVTSDL-----DPTQVILPLKTLNVA-----SVPIYMSW 113  
 79 CG--SDLANGIGERDVSVGEENLATGTLASSGATARTIVRYKPLVYKIPYTTW 136  
 114 SPLQON-FWVEDEVTL--HNIPY--MGDEVL---DQGTPIEB-----LKNYDGKV-- 157  
 137 IFLDKKQRMADQSVGRRRIYDTVGNBALICSDDEBEIPEPEEKHFFTKGEDHLWR 196  
 158 ----HGDRCGFINDIPIVELYNALGQYNDDDDDGDDPPEEREKQDLJEDHRDKESR 213  
 197 ATQDGG-----LNGBV-----VNLCOF-----IGATPSBIERSSEVLFR-KNEKISG 238  
 214 PPRKPPS---DKIFPAISMPFDKGTABELKRYKELTEQQLPGALPPECTPINDGPA 269  
 239 SSDKIESRLSLDKTMADV----- 256  
 270 KSVQBOSLHSPHLLFCRRCFKYDCFLH-----PFAHPNTYKRNKTETALD 316  
 257 ----LDSFDNLFCCRCLVDFCRHGCSONLVPRCEKQYSPDPENK----- 299  
 317 NKPCPGCY---QHLEKKE----- 333  
 300 -FPCGHLCYLRPPQWMBGFKEMHMDGLAGATYTESGASQAVDNNVYSESDSNRKG 358  
 334 ----PAAATTAERIKTP-----KRPGRRR 355  
 359 NTRSMVLVGTSGSKIISVASESTTTPADISETNVSDLPSSLRKHKISKIGPRRR 418  
 356 GRLPNNSRRPTPTI-----NVLESKDTSDREAGTETGENDKEBEKKDETS 405  
 419 BHSFGKQKVTSDISFEGNIMNKLSTPEIRDT--RLSRRSGG-----DLRLIDEST 470  
 406 SSSSEANSRC-QTPIRKKNIEPPEN-----VR-----NSGAERSMFRVLIGTYDNF 451  
 471 KTSRKMDGMBPATYTMENVGRQSNKVYSTKNFLSTLSCWALERDLYLKQIRIFGKNS 530  
 452 CALIAR--LIGTTCQOYVEFRVKESSIIAPAP-----ADVDVTPPK 491  
 531 CLIANRLSLGLTCTEVANVTYNNGAAMAKRPLANKSISGDPAENQDYMEDDMARTTI 590  
 492 KKKRKR-----LW--AAHCRKIQLKDGSSNMHVNYOPCDHPROPCCSSCPVIAONFC 543  
 591 YRRGRNRRLKYTWASAGH-PTVRRKRTDQKQCTYQYSPC-ACQMGCGMDCCADKGTCC 648  
 544 ERFCCSSBECQNRFPQRCR-KAQCNTKQCPCTLAVERCDPDLCTLGAADHWS----- 596  
 649 EKYCCGSSCKNRFRCCHCAKQSCRSRQPCFPAASRECDPVCRCN-----WVSCGDSGL 703  
 597 ----KNVSCNCSIORGSKKHLIAPSDVAGWGIPTKDPQXKKEPISSEKGLIISOD 649  
 704 GEPPLARGDGYQCENMGLLKQOQRILLGSDVAGWGAFTKPNVNRNDYLGSTTGBLISHK 763  
 650 EADRRGKYVDKYMCSFLFNINLNDPVVDATRKGNKIRFANHSVNPNCYATV 699  
 764 EADKRGKITYDRANSSFLFDLNDQYVLDAYRKGDKLKFANHSSNPNCYATV 813

RESULT 8  
 US-11-330-403-9237  
 Sequence 9237; Application US/11330403  
 Publication No. US20060159563A1  
 GENERAL INFORMATION:  
 APPLICANT: Abad, Mark S.  
 TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 FILE REFERENCE: 38-21(53629)B  
 CURRENT APPLICATION NUMBER: US/11/330,403  
 CURRENT FILING DATE: 2006-01-12  
 NUMBER OF SEQ ID NOS: 19250  
 SEQ ID NO 9237  
 LENGTH: 895  
 TYPE: PRT  
 ORGANISM: Oryza sativa (indica cultivar-group)  
 US-11-330-403-9237

Query Match 17.3%; Score 658; DB 7; Length 895;  
 Best Local Similarity 25.9%; Pred. No. 8.8e-33;  
 Matches 200; Conservative 89; Mismatches 235; Indels 248; Gaps 26;

95 VPLKTLNVAASVPIMYSPLQON-FWVEDEVTLHNIPYMGDEVLDQGTPIEBILKNY 153  
 144 VQJLAKLPVVDKIPYTTWIFLDKNQRMADQVLCRRRIY-----Y 184  
 154 DGVHGDRCGFINDIPIVELYNALGQYNDDDDDGDDPPEERE-----EKQ---KOLE 204  
 185 PIVNEALIC-----SSDDVDPEPEEREKVFTEGEBQILWKATQ 224  
 205 DHRDKESRPPKPPSDKIFPAISMPFDKGTABELKRYKELTEQQLPGALPPECTPNI 264  
 225 DHGLSR-----VNLVLCQFVD-ATPSBIERSSEVLFRKY--EKQSSSYET 268  
 265 DGNPAKSVQBOSLHSPHLLFCRRCFKYDCFLH-----PFAHPNTYKRNKTETALDNK 318  
 269 DQQLFLGTMVDALDSFNLFCRRCLVDFCRHGCSONLVPRSEKQPIGHGDE---NKR 325  
 319 PCGPQCY---QHLEKKE-----QHLEKAPPAALTAERIKTPPKRGRRR 355  
 336 PCGDQRYLRREVVYODTCNDRNACTYNTDSRSSSLKVSATILSR----- 371  
 356 GRLPNNSRR--PSTPIVNLBSKDTSD----- 381  
 372 ---SEDSNRDEDNITSTIVTSRKTITNSRYADKSVTPPGDASETNVSPDMLRTYG 428  
 382 REAGTETGENDKEBEKKDETS-----SEANSRC---QTPIRKKNIE 424  
 429 RRTSKHASKNDHSPDRQKIYSSPPFPAASVANKQSVPRIGETCPDISBAVDQLRSL 488  
 425 BPP-----ENV-----EWSGAERSMFRVLIGTY 448  
 489 DDPNKISTKMDGMBPATYTMENVGRQSNKVYSTKNFLSTLSCWALERDLYLKQIRIFG 548  
 449 DNFCALIAR--LIGTTCQOYVEFRVKESSIIAPAP-----ADVDVTP 488  
 549 KNSCLIANRLSLGLTCTEVANVTYNNGAAMAKRPLANKSISGDPAENQDYMEDDLVAR 608  
 489 PRKKRKRKR-----LMAHCRKIQLKDGSSNMHVNYOPCDHPROPCCSSCPVIAONFC 541  
 609 TRICRRKRAKRLKYTWASAGHPYRRKIRIGQKQYTYQYNPCG-CQMGCGMDCAVENGT 667  
 542 FCFECCSSBECQNRFPQRCR-KAQCNTKQCPCTLAVERCDPDLCTLGAADHWS----- 596  
 668 CCEKYCCGSSCKNRFRCCHCAKQSCRSRQPCFPAASRECDPVCRCN-----WVSCGDSGL 722  
 597 ----KNVSCNCSIORGSKKHLIAPSDVAGWGIPTKDPQXKKEPISSEKGLIISOD 647  
 723 SLGEPPLARGDGYQCENMGLLKQOQRILLGSDVAGWGAFTKPNVNRNDYLGSTTGBLISHK 782  
 648 QDEADRRGKYVDKYMCSFLFNINLNDPVVDATRKGNKIRFANHSVNPNCYATV 699  
 783 HREADKRGKITYDRANSSFLFDLNDQYVLDAYRKGDKLKFANHSSNPNCYATV 834



QY 342 RIKTEPK--RPGRRRGRRLPNNSRSTPTINVLBSKOTDSDREAGTETGENNDKEE 399  
DB 472 PPLPPRLRP-----PSPPPEETP-----EPKPPVPLAPPEDDHP 509  
QY 400 KQDETSSSEANSRCOTPIKMKPNIEPPENVEGABAMFVLTGTYDNCALRLIG 459  
DB 510 PRTPLGSLAKSGSTETVPATPGGEPPLSGSGSLSPQLPFEELDNQMPSEALIPG 569  
QY 460 TTTCTQVVEP-----RVKSSIIAPAEVDVTPPKKKK-----494  
DB 570 PRDEVTETVVLAKVGPWRBPPKKHEDLVAPS---DESPPOPLFRPSGFSEMTIL 626  
QY 495 -----KRLMAHCKKIQLK-----DGSNNH- 517  
DB 627 YIIMNGGIDBEDIRLCVTERLLQDNGMDMLNDTLWYHPILPDKKREDEGREHVT 686  
QY 518 -----YNOPCDHP-----QPCDSCPCYIAONPCEKFCQSCSECONR-- 556  
DB 687 GCABSEGFYTIIDKDKLRLNASSRSTDEPMDTQMSIFPAQPHASTRAGSRSE-QRLL 745  
QY 557 --PQGRCAQCNTRKQPCYLAVRBCDPLCLTGAADHMSKNVSCNKSIOGSS--K 611  
DB 746 SSFTG-----SCSDLL-----KFNQLKVLRLDLRGSQFRK 776  
QY 612 KHLIAPSDVAGWGIPIKDPVOKNEPISYCGEIIISQDEADRRGRYDK--YMCSTLPTL 669  
DB 777 KLLCKCKSHIDWGLPAMEPIADEWIVYVQNIQVLAIDMKREKKEVIGIGSSYMFV 836  
QY 670 NNDPVVDATRKGNKIRFANHSVNPNCYAV 699  
DB 837 DHDITIIDATCKGNFARFINHSCNPNCTAV 866

RESULT 13  
US-11-330-403-16696  
; Sequence 16696, Application US/11330403  
; Publication No. US20060159563A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53629)B  
; CURRENT APPLICATION NUMBER: US/11/330,403  
; CURRENT FILING DATE: 2006-01-12  
; NUMBER OF SEQ ID NOS: 19250  
; SEQ ID NO 16696  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-11-330-403-16696

Query Match 5.7%; Score 218; DB 7; Length 497;  
Best Local Similarity 26.0%; Pred. No. 4.8e-06;  
Matches 53; Conservative 25; Mismatches 58; Indels 68; Gaps 5;

QY 502 HRRK-----IQKKQSSNNHYNYQPCDHPQPCDSSCPVIAQNFBEKFCQSSBQNF 557  
DB 266 HIRNRIYLVKKRQDANDEV---GCTNCGPDCDSVCGRVOCISCKGCSPEBQNR- 320  
QY 558 PCRCRCAQCNTRKQPCYLAVRBCDPLCLTGAADHMSKNVSCNKSIOGSSKKHLLIA 617  
DB 321 -----PRKREKKIKIV 331  
QY 618 PSDVAGWGIPIKDPVOKNEPISYCGEIIISQDEADRR-----GKYVDKYMCSPLFNILN 671  
DB 332 KTEHCGWGYBAASINKEDFIVEYIGEVISDAQCGRLMDMKHKKMDPYMC-----EIOK 387  
QY 672 DTVVDATRKGNKIRFANHSVNPNCY 695  
DB 388 DFTIDATPKGNASRFLNHSNPNPC 411

RESULT 14

US-11-330-403-2871  
; Sequence 2871, Application US/11330403  
; Publication No. US20060159563A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53629)B  
; CURRENT APPLICATION NUMBER: US/11/330,403  
; CURRENT FILING DATE: 2006-01-12  
; NUMBER OF SEQ ID NOS: 19250  
; SEQ ID NO 2871  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: ASPERGILLUS NIDULANS FGSC A4  
US-11-330-403-2871

Query Match 5.7%; Score 218; DB 7; Length 980;  
Best Local Similarity 21.3%; Pred. No. 1.1e-05;  
Matches 87; Conservative 63; Mismatches 138; Indels 120; Gaps 18;

QY 325 YQHEGAKFPA-ALTARIK-----TPPKRPGRRGRRLPNNSRSTPTI----- 370  
DB 4 HMDADROSEFPVADAVTAMKLEQSENNTDAPILNGG--AMKPDKASPEPLKDERAS 61  
QY 371 -NVLESKOTDSDREAGTETGENNDK--EBEKQDETSSSEANSRCOTPIKMKPNIEBP 427  
DB 62 STFMKSRSSRTSPSSRTPLKKEHSDSDIOEKRGDASGTEKVGGIS--VAMEBG-QPP 118  
QY 428 ENVEGABAMFVLTGTYDNCALRLIGTKTCROYVEFRVKESSIIAPAPAE-DVD 486  
DB 119 K-----LAR-----SSSQVVRPQLPLFD 138  
QY 487 TPKKKKKRRLMAHCKKIQLKQDSSNNHYNYQPCDHPQPCDSCPCVY----- 538  
DB 139 LPDSTBAQKTEVEVLETCQYANKWGYTEHA---NECD-----CAEHWLVVLVLAAPSPF 190  
QY 539 ---AQN-----PCEKFCQSCSECONRPPGRCRCAQCNTRKQPCYLAVRBCDPLCLTGAAD 592  
DB 191 RVPSONPASSTNRACGSDSDCLNRYTKIECGMDCG-----GQPD----- 229  
QY 593 HWDKNSVCKNKSIOGSKKHLIAPSDVAGWGIPIKDPVOKNEPISYCGEIIISQDEAD 652  
DB 220 -----QONRPPRRRYAVAVIKTEKKQGLAABEDLRHQCIFVEYGEVINEGPPH 281  
QY 653 RRGKYVD---KYMCSPLFNILNDFVVDATRKGNKIRFANHSVNPNCY 696  
DB 282 RMRQYDAKGIKHF--YFMSSKGFVDATKKNIGARFCNHSNPNCTAV 327

RESULT 15  
US-10-539-228-358  
; Sequence 358, Application US/10539228  
; Publication No. US20060154250A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris

; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: CHIR0052-101 (PP023370,0003)  
; CURRENT APPLICATION NUMBER: US/10/539,228  
; CURRENT FILING DATE: 2005-06-17  
; PRIOR APPLICATION NUMBER: US 10/322,281  
; PRIOR FILING DATE: 2002-12-18  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 358  
; LENGTH: 870  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-539-228-358

Query Match 5.5%; Score 210.5; DB 6; Length 870;  
Best Local Similarity 28.5%; Pred. No. 2.7e-05;  
Matches 65; Conservative 22; Mismatches 78; Indels 63; Gaps 7;

```
QY 498 LMAAHCRIOLKKOSSNHV-----YNYQPCDHR-----QPCDSCPCV 537
Db 630 LMTHTPSKKKRDGDIRHVTGCASSBGFYTIIDKKDKLRYNSSRASTDEPPADTQMSI 689
QY 538 IAQNFCEKFCOCCSSSECQNR---FPGCRCKAQCNTPKQPCYLAVERCDPDLCLTGAADH 593
Db 690 PAQPHASTRAGSRSE-QRRLSSFTG-----SCDSDL----- 721
QY 594 WDSKNVSCNCSIQKGSCKHLLAPSDVAGWCIPIKDPVQKNRPISEYCGEIIISQDEADR 653
Db 722 -----KFNQLKFRKKQKFKCKSHIDWGLFAMEPIADEWVLETVGQNIHQVIADM 772
QY 654 RGKYYDK--YMCSPLEFNANDFVVDATRKGNKIRFANHSVNPNCYAKV 699
Db 773 REKRYEDEGIGSYMFRVHDITIIDATKCGNPARFINHSCNPNCYAKV 820
```

Search completed: August 11, 2006, 22:50:49  
Job time : 25.7682 secs

Copyright (c) 1993 - 2006 Bioacceleration Ltd.

Run on: August 13, 2006, 13:11:01 | Search time 1186.4 Seconds  
(without alignments)  
6396.808 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGQYKSEKBPVCKRRKRVK.....RYSQADALKYGIENEMRIP 746

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODBI=frame+ p2n.model -DEV=nlh  
-O=/abs/ABSSWB spool/US10773302/runat.11082006.140223.16618/app.query.fasta\_1  
-DB=GenEmbl -QFMT=fastap -SUPPLX=p2n.rge -MINMATCH=0.1 -LOOPCC=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs804  
-USER=US10773302 @CCN 1.1.13924 @runat.11082006.140223.16618 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DEBLOCK=100 -LONGLOG -DEF TIMEOUT=120  
-WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl.\*

1: gb env:\*  
2: gb pat:\*  
3: gb ph:\*  
4: gb pl:\*  
5: gb pr:\*  
6: gb ro:\*  
7: gb str:\*  
8: gb sv:\*  
9: gb un:\*  
10: gb vl:\*  
11: gb ov:\*  
12: gb hcg:\*  
13: gb in:\*  
14: gb om:\*  
15: gb ba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	4051	100.0	2511	2 CQ714720
2	4051	100.0	2600	2 A58329
3	4051	100.0	2600	2 AR473734

4	4051	100.0	2600	2 AX33675	AX33675 Sequence
5	4051	100.0	2600	5 HG61145	U61145 Human enhan
6	4046	99.9	2619	5 AB179385	AB179385 Macaca fa
7	4038	99.7	2558	2 BD186247	BD186247 STAT act
8	4038	99.7	2576	2 AX821950	AX821950 Sequence
9	4038	99.7	2576	2 HSZESTENH	X95653 H.sapiens m
10	4030.5	99.5	2655	5 BC010858	BC010858 Homo sapi
11	4024	99.3	2512	2 DD208558	DD208558 Expressio
12	3999	98.7	2574	6 BC003772	BC003772 Mus muscu
13	3999	98.7	2595	6 BC016391	BC016391 Mus muscu
14	3999	98.7	2600	6 AK220174	AK220174 Mus muscu
15	3993	98.6	2653	6 MMU52951	US2951 Mus muscu
16	3922	96.8	2617	6 BC079538	BC079538 Mus muscu
17	3829	94.5	2631	11 CR855647	CR855647 Xenopus t
18	3812.5	94.1	2476	2 AR118692	AR118692 Sequence
19	3812.5	94.1	2476	2 AR747667	AR747667 Sequence
20	3812.5	94.1	2476	5 AK092676	AK092676 Homo sapi
21	3809	94.0	2658	11 BC084193	BC084193 Xenopus 1
22	3793	93.6	2772	11 AP351126	AP351126 Xenopus 1
23	3783.5	93.4	3641	5 AB208895	AB208895 Homo sapi
24	3758	92.8	2636	11 BC097526	BC097526 Xenopus 1
25	3698.5	91.3	2546	5 AB168941	AB168941 Macaca fa
26	3369	83.2	3394	11 AB195561	AB195561 Oryzias 1
27	3362	83.0	2666	5 HSU52965	US2965 Human putat
28	3106	76.7	1960	2 BD186308	BD186308 STAT6 act
29	2770	68.4	4323	11 TMI313481	AJ313481 Tetracodon
30	2606.5	64.3	2253	6 AF483490	AF483490 Mus muscu
31	2606.5	64.3	2253	6 AF483491	AF483491 Mus muscu
32	2606.5	64.3	3970	6 MMU60453	U60453 Mus muscu
33	2606.5	64.3	4179	6 BC007135	BC007135 Mus muscu
34	2606.5	64.3	4185	6 AB004817	AB004817 Mus muscu
35	2603.5	64.3	2244	8 BT009782	BT009782 Homo sapi
36	2603.5	64.3	2244	8 AT889044	AT889044 Synthetic
37	2603.5	64.3	2594	4 CQ715926	CQ715926 Sequence
38	2603.5	64.3	4606	5 AB002386	AB002386 Homo sapi
39	2603.5	64.3	4640	5 BC015882	BC015882 Homo sapi
40	2601.5	64.2	2594	2 AX821949	AX821949 Sequence
41	2601.5	64.2	2594	2 HSU50315	US50315 Human enhan
42	2581	63.7	4639	5 HSM80470	AK33159 Homo sapi
43	2579.5	61.7	4147	6 AK129004	AK129004 Mus muscu
44	2306	56.9	2749	5 AB004818	AB004818 Homo sapi
45	2156.5	53.2	2439	13 AY051785	AY051785 Drosophila

## ALIGNMENTS

RESULT 1  
LOCUS CQ714720 2511 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 654 from Patent WO02068579.  
ACCESSION CQ714720  
VERSION CQ714720.1 GI:42275577

KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS  
TITLE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

JOURNAL  
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
Kites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
Patent: WO 02068579-A 654 06-SEP-2002;  
PE Corporation (NY) (US)

FEATURES  
source  
location/Qualifiers  
1..2511

ORIGIN  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

Alignment Scores:



Pred. No.: 0 Length: 2511  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0

US-10-773-302-2 (1-746) x CQ714720 (1-2511)

QY 1 Metcylglnthrnglylylvalyserserglulysgliprovallcysrtparqlysarvallys 20  
DB 10 ATGGGCCAGACTGGGAGAAATCTGAGAGGAGCAGGTTGTGGCGAAGCGGTAAAA 69  
QY 21 Serqlutymeacargleuargluleuylsarpheargalaaapgluvallyseser 40  
DB 70 TCAGAGTACATGACACTGAGCAGCTCAAGAGTTTCAGACGAGCTGAAGTAAAGGT 129  
QY 41 Metpheserseraanaarglulyslleuqluarqthrglulleuamrglulgutp 60  
DB 130 ATGTTAGTTCCAATCGTCAGAAAATTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 189  
QY 61 Lyseglnarqargllelnprovalhlslleuthrservaliseserserleuarglythr 80  
DB 190 AAACAGCGAAGATACAGCTGTGCAATCTGACTTGTGAGCTCATTCGCGGAGCT 249  
QY 81 Argglucyservalthrseraspleuapheprothrlnvalilleproleuylthr 100  
DB 250 AGGAGTGTTGGTGACCAAGTCACTTGATTTTCCAAACAAAGTCATCCATTAAAGACT 309  
QY 101 Leuamnalavalalaservalprolleuyltysertpserproleuqlnlnasphe 120  
DB 310 CTGAATGCAATGCTTCAGTACCCATATGTAATCTTGCTCCCTACACAGAAATTTT 369  
QY 121 Metvalgluaapgluthrvalleuhsanilleprotyrmetgluasrgluvalleuap 140  
DB 370 ATGGGGAGAGTGAACCTGTTTACATACATCTCTTATATGGAGATGAAGTTTATAT 429  
QY 141 Glnasrglythrphelleugluleuulleysanlyzaasrglyllysvalhlsagllyap 160  
DB 430 CAGAGTGTACTTTTCATTGAAGAACTAATTAATAATTTATGATGGAAAGTACACGGGAT 489  
QY 161 Argglucysglipheulleasnaapglullephevalgluleuvalasnaalleuqln 180  
DB 490 AGAGATGTGGGTTATATAATGATGAATTTTGTGAGATGGTGAAGCCCTTGCTAA 549  
QY 181 Tyrsnaapraapraapraapraapraapraapraapraapraapraapraapraap 200  
DB 550 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 609  
QY 201 Lysaspleuqluasrphlsarqasrphlysgluserarqproproarglyllyspheproser 220  
DB 610 AAAGATCTGGAGGATCACCGAGATGATTAAGAAACCGCCACCTCGAAATTTCTCT 669  
QY 221 Asplysllepheglualalleesersemetphepoadrlyllythrlalglulgu 240  
DB 670 GATATAATTTTGAAGCAATTTCTCAATGTTTCCAGTAAAGGACACGAGAAACATA 729  
QY 241 Lysglulyslylzeugluleuthrglulnglulneuproglylalaaleuoproglucys 260  
DB 730 AAGGAAAATATTAAGAACTCACCGAACGAGCTCCAGGCGACTTCTCTGAAATGT 789  
QY 261 Thrproaenlleasrglylproaenlalysservalglnatrglulnserleuhseser 280  
DB 790 ACCCCCAACATGAGTGAACAATCTTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 849  
QY 281 Phehlsrthrleuphecyatarglycyspheylslyzarpqysrphleuhsisprophehs 300  
DB 850 TTTCAATAGCGCTTCTGTAGCGATGTTTAAATGACTGTCTTCTCAATCCCTTTTAT 909  
QY 301 AlaThrproaenThrlyrlylsarqlylsanthrqlutrralaaleuapraapraapraap 320  
DB 910 GCACACCCCAACATTTATAGCGAGAGAACACAGAAACAGCTTGAACAAACCTTGT 969

QY 321 Glyproqlncyslyrqlnhlsleuqlulglvalalysglulphealaaalaleuthrlla 340  
DB 970 GGAACCAAGGTATACAGCAATTTGGAGGGAGCAAGAGATTTGTGTGCTTCACCGCT 1029  
QY 341 Gluabrglleysrthrproprolylsarqprogllyllyrargrarglylzeuuproasn 360  
DB 1030 GAGCGGATTAAGAACCCCAACAAACCTTCAGAGAGCCGCGAGAAAGACCGCTTCCAT 1089  
QY 361 Asnserseraqproserthrprothrllaasnaalleuqluserlysaapthrapseser 380  
DB 1090 AACAGTAGAGGCCACAGACCCCAACATTATGTCTGAATCAAGATATACAGACGT 1149  
QY 381 Asparglualaglythrqlutthrqllyglulbasnaapraapraapraapraapraap 400  
DB 1150 GATGGGAGACAGGACATGAACCGGGGGAGAGAACATATTAAGAAAGAAAGAAAG 1209  
QY 401 Lysasrgluthrserserserserserglualaasnsarqysglnthrproilleuamet 420  
DB 1210 AAAGATGAACCTTGAGCTCTCTGAAAGCAATTTCTGGGTCTCAACACCAATTAAGATG 1269  
QY 421 Lysproaenlleuoproprogluasnavalglutpserglvalagluasermetphe 440  
DB 1270 AAGCCAAATATTGAACCTCTGAGAAATGTGAAGTGAAGTGTGCTGAAGCTCATGT 1329  
QY 441 Argvalleulleuqllythrlyrlyrarsapraaphecyalallealargleulleuqlthr 460  
DB 1330 AGAGTCTCATTTGCACTTACTATGACAAATTTCTGTGCCATGTGATGTTAATGGACC 1389  
QY 461 LysrthrCyasrarglnalysrgluPheargvallysgluserserllellealproala 480  
DB 1390 AAAACATGTACAGAGGTATGAGTTAGAGTCAAAATACTTAGCATCATAGCTCCAGCT 1449  
QY 481 Proalagluasrvalaapthrproproarglyllylsarqlyshlsarqleutrrala 500  
DB 1450 CCGGCTGAGATGAGATACCTCTCAAGAAAGAAAGAAAGAAACCGGTTGTGGGCT 1509  
QY 501 AlahlsacyasrarglylleuqlneuleuylasapglyseserseranhlsvallyrAsnlyr 520  
DB 1510 GCACATGCAAGAAAGATACAGCTGAAGAAAGACGCGCTCTTAAACATGTTTCAACATAT 1569  
QY 521 GlnproCyasrphlsproarglnproCyasrpserserCyproCyasvalillealagln 540  
DB 1570 CAACCTGTATATCAACCGGAGCTTTGACAGTTCTGTGCTTGTGATGACACAA 1629  
QY 541 AsnphesysglulyspheCyeglnCysserserglucysgluasnaarqheprogllyCys 560  
DB 1630 AATTTTGTGAAAAGTTTGTCAATGTATGATGATGATGATGATGATGATGATGATG 1689  
QY 561 ArgcyalsyalaglnCyasenthrlyaglnCyserprocyerlyrleualavalargglucys 580  
DB 1690 CGCTGCAAGACAGTCAACACCAACAGAGCGCGGTCTACCTGGCTGTCCGAGAGTGT 1749  
QY 581 AspproaspleuCybleuthrCyeglylalaasrphlsrtpaserserlysaasnaalyseser 600  
DB 1750 GACCTTACCTCTGTCTTACTTGTGAGCGGCTCAACATTTGGGACATGAATAAATGTGTCC 1809  
QY 601 CyslysaenCyserillelnatrglylserlylsylsleuleuleuallaaproserap 620  
DB 1810 TGCAAGAACTGCAATATTACAGCGGGGCTCCAAAAGATGATGATGCGGACCAATCGAC 1869  
QY 621 Valalaglyrtrpglyllepheulleuapraprovalglulysaanglupheilleseglu 640  
DB 1870 GTGGCAGCGTGGGGAATTTTATGAAGATCTGTGCAAGAAATGAATTCATCTCGAAA 1929  
QY 641 Tyrcysglulullelleserglnasrgluallaasrargrarglyllysvallyzarplys 660  
DB 1930 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTACAGAGAGGAAAGTATAGATAAA 1989  
QY 661 TyrmetCyssersrphleuapheasnaalleuapraapraapraapraapraapraapraap 680  
DB 1990 TACATGTGAGACTTCTGTTCATCTGAACATGATTTTGTGAGTGAACCCGCGAAG 2049  
QY 681 Glyasnaalleasrphelleasnaalleasnaalleasnaalleasnaalleasnaalleasna 700

Db 2050 GGTAAACAAATTCGTTTGCAATCACTCGTAATCCAACTGCTATGCAAAATTTATG 2109  
Qy 701 MetValaengIyaSPHisArgIleGIlyIlePhealalyaArgAlaIleGIInThrgIyGIu 720  
Db 2110 ATGGTTAAACGGATGACAGCATAGTATTTTGGCAAGAGCCATCCAGCATGGCGGA 2169  
Qy 721 GtUleuphnePheAspTyrArgTyrSerGIInAlaAspAlaLeuIlyTyrValGIlyIleGIu 740  
Db 2170 GAGCTGTTTTTGTGATTACAGATACAGCCAGCTGATGCCCTGAAGATAGTCGGCATCGAA 2229  
Qy 741 ArgIuMetGIuIlePro 746  
Db 2230 AGAGAAATGGAATCCCT 2247  
RESULT 2  
A58329 2600 bp DNA linear PAT 05-MAR-1998  
LOCUS Sequence 1 from Patent WO9635784.  
ACCESSION A58329  
VERSION A58329.1 GI:3713991  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1  
AUTHORS Jenuwein T. and Laible G.  
TITLE CHROMATIN-REGULATOR GENES  
JOURNAL Patent: WO 9635784-A 1 14-NOV-1996;  
BOEHRINGER INGELHEIM INT (DE)  
COMMENT Other publication DB 19516776 961114.  
FEATURES  
source 1. 2600  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/cell\_type="B-CELL"  
1. 89  
90. 2330  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAA03482.1"  
/db\_xref="GI:3713992"  
/translation="MGQTGKSEKGPVCVRKKVSEYMLRQLKFRPADDEVKSMFSS  
NRKTLBRTILNOMKORIQPVHILTSVSLRGTRBGSVSDLPFOVLPKTLN  
AVASVPIWMSNPLOQNPVDEBTYLNHPVMDGVLDODGPTIELIKNDGKVGHD  
RECGFINDEIFVELNALGIYNDDDDDDDGDDPERBERKQDLKEDHDDKSRPRKX  
PSDKIFEAISNFPDKGTAEELKEKRYELTEQLPGALPPECTPNIIDGNASKVREQ  
SLASHFTLFCRCRCFYDCFLHPFATPTPKNTETLNDPCGPOCCHLEGAKF  
AAALTAERIKTPKRGGRGRRLPNNSRSPSTINVLSESDTDSBAGETGSEN  
NDKEEEKKDETSSEBANSRCOTPIKKKPNTEPPENYEMSGAASMTFVILGTYDN  
PCATNRLIGTKTCROYTERVRESSTIAPAPAEVDYTPPRKKRGRHMAAKRTIOL  
KDDSSNHYNTQPCDHPQPCDSSCPCTAQNFCERKQCSCSECONRPGCKCAQC  
NFKQPCYLAVERCDPDLCTCGAADHMDKNSCKNCSIQHSKHLILVADVAG  
GIFIDPOVKNFISEYCEGILISODEADBRGVYDKMCSFLFNINDFVADATKGN  
KIRFANHSVNPICVAKVMVWNGDHRIIGIFAKRAIGTEBELFDYRYSDALKVYIGIE  
REMEIP"  
ORIGIN 3'UTR 2331. 2600  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.08  
Best Local Similarity: 100.08  
Query Match: 100.08  
DB: 2 Gaps: 0  
US-10-773-302-2 (1-746) x A58329 (1-2600)  
Qy 1 MetGIyGIInThrgIlyIlyIlySergIlyIlyGIyProValCyeTPrArgIlyArgValIys 20

Db 90 ATGGCCGACGATCGGGAAGAAATCTGAGAGGAGCACAGTTTGTGGCGGAGGTATAAA 149  
Qy 21 SergIlyrMetArGleuArgIleuIlyArgPheArgAlaIleAspGIuValIlySer 40  
Db 150 TCAGAGTACATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 209  
Qy 41 MetPheSerSeraMaArgIleuIlyIleuGIuArgTyrThrgIuIleuAspGIuIlyr 60  
Db 210 ATGTTAGTTCGAAATCGTCAGAAATTTTGGAAAGAACGAAATCTTAAACAGAAATG 269  
Qy 61 LysGIuArgArgIleGIInProValHisIleuThrSerValSerSerIleuArgGIyThr 80  
Db 270 AAACAGCAAGATGACAGCTGACATCTGACATCTGAGAGTCAATTCGGCGGACAT 329  
Qy 81 ArgGIuCySerValThrSerAspIleuAspPheProThrgIuValIleProLeuIlyThr 100  
Db 330 AGGAGGTGTCGATGACATGACATGATGATGATGATGATGATGATGATGATGATGAT 389  
Qy 101 LeuAsnAlaValAlaSerValProIleuMetTyrSerTyrSerProLeuGIuIleuAsp 120  
Db 390 CTGAATGCAATGCTTCAGTACCAATTAATGATTCCTGCTCCCTACAGCAATTT 449  
Qy 121 MetValGIuAspGIuThrValIleuHisAsnIleProTyrMetGIyAspGIuValIleuAsp 140  
Db 450 ATGTTGAGAGATGAACTGTTTACATTAACATCTTATATGAGAGATGAACTTTAGAT 509  
Qy 141 GIuAspGIyThrPheIleGIuIleuIleIlyAsnTyrAspGIyIlyValHisGIyAsp 160  
Db 510 CAGATGTGATCTTCATTAAGAACTAATAAAATTAATGATGAGAAATGATCACCGGGAT 569  
Qy 161 ArgGIuCySerGIyPheIleAsnAspGIuIlePheValGIuIleuValAsnAlaIleuGIu 180  
Db 570 AGAGATGTGGGTATTAATGAATGAATTTTGTGGAGTGGTAATGCTTGTGTCTCA 629  
Qy 181 TyrAsnAspAspAspAspAspAspGIyAspAspProGIuGIuArgGIuIlyGIuGIu 200  
Db 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
Qy 201 LysAspIleuGIuAspHisArgAspAspIlyGIuSerArgProProArgIlyPheProSer 220  
Db 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
Qy 221 AspIlyIlePheGIuAlaIleSerSerMetPheProAspIlyGIyThrAlaGIuIleu 240  
Db 750 GATTAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGCAAAACATA 809  
Qy 241 LysGIuIlyrGIyIlyGIuIleuThrgIuGIuGIuIleuProGIyAlaIleuProProGIu 260  
Db 810 AAGGAAATATTAAGAACTCACCGAAGACGACTCCAGGCGCACTTCTCTGAATGT 869  
Qy 261 ThrProAsnIleAspGIyProAsnAlaIlySerValGIuArgGIuIleuIlySer 280  
Db 870 ACCCCCAATATAGAGAGCAAAAGCTAATCTGTTGAGAGAGCAAAAGCTTACATCC 929  
Qy 281 PheHisThrIleuPheCysArgArgCysPheIlyrGIyAspCysPheIleuHisProPheHis 300  
Db 930 TTTCTATAGCTTTCTCTGTAAGCGAATTTTAATGTGATGCTTCTTCAATCCCTTTTAT 989  
Qy 301 AlaThrProAsnThrTyrIlyAspIlyAspAsnThrgIuThrAlaIleuAspAsnIlyProCys 320  
Db 990 GGAACACCAACATTAATAG 1049  
Qy 321 GIyProGIuIlyrGIyIleuIleuGIuIlyAlaIlyGIuIleuPheAlaIleuThrAla 340  
Db 1050 GGAACACAGTGTATACATGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109  
Qy 341 GIuArgIleuIlyrThrProProIlyArgProGIyGIyAlaArgArgArgIlyArgIleuProAsn 360  
Db 1110 GAGCGAGATTAAGACCCCAACAAACGTCCAGAGAGCGCAGAGAGAGAGAGAGAGAG 1169  
Qy 361 AsnSerSerArgProSerThrProThrIleuAsnValleuGIuSerIlyAspThrAspSer 380

Db 1170 AACAGTAGCAGGCCAGCAGCCCAACATTAATGTGCTGAATCAAGATACAGACGT 1229  
Qy 381 AApArpglualaglythrGluThrGlyGlyGluAsnAsnAAspLySGluGluGly 400  
Db 1230 GATAGGAGAGCAGGAGCTGAAACGGGGGAGAGAACTATGATTAAGAGAGAGAGAG 1289  
Qy 401 LysApGluThrSerSerSerSerSerGluAlaAsnSerArgGlnThrProIleuMet 420  
Db 1290 AAAGATGAAACCTTGAGCTCTCTGAAACCAATCTCGGTGTCAACCAATTAAGATG 1349  
Qy 421 LysProAsnIleGluProProGluAsnValGluThrSerArgAlaGluAlaSerMetPhe 440  
Db 1350 AAGCAATATATGAACCTCTGAGAAATGTGAAGTGAAGTGAAGTGAAGTGAAGT 1409  
Qy 441 ArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1410 AGAGTCTCATTTGGCACTTACTATGACAAATTTCTGTGCCATTTGCTAGTTAATGGAC 1469  
Qy 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleAlaProAla 480  
Db 1470 AAAACATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGCATCATAGCTCAGCT 1529  
Qy 481 ProAlaGluAspValAspThrProProArgLysLysLysLysLysLysLysLysLys 500  
Db 1530 CCGGTGAGATGTGGATATCTCTCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1589  
Qy 501 AlaHisCysArgLysIleGluLeuLysAspGlySerSerAsnHisValTyrAsnTyr 520  
Db 1590 GCACCTGCAAGAAAGATCAAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649  
Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1650 CAACCTGTGATCATCCAGCGAGAGCTGTGACAGATGTGTCCTTGTGTATGACAA 1709  
Qy 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyLys 560  
Db 1710 AATTTTGTGAAGATTTTGTCAATGATGTCAAGGTGCAAAGCCGCTTCCGGAGAGC 1769  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys 580  
Db 1770 CGCTGCAAGCAGAGTGCACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1829  
Qy 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
Db 1830 GACCTGACCTCTGTACTTACTGTGAGCGCGCTGACCATTTGGGACAGTAAATGTGCC 1889  
Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysLysLysLysLysLysLysLys 620  
Db 1890 TGCAAGAACTGCAGATTCAGCGGGGCTCCAAAGAGCATTAATGCTGGCACATCTGAC 1949  
Qy 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 1950 GTGGAGGCTGGGGAGATTTTATCAAAAGATCTGTGCGAGAAAGATTAATCTCAGAA 2009  
Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAGGGAGAAAGTGTATGATA 2069  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
Db 2070 TACAGTGCAGCTTCTGTCTTCAACTTGAACATGATTTGTGTGATGCAACCGCGAAG 2129  
Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
Db 2130 GGTAAACAAATTCGTTTTCGCAATCATTCGTAATTCGTAATTCGTAATTCGTAAT 2189  
Qy 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
Db 2190 ATGGTTAACGGTATCACAGATAGTATTTTGGCAAGAGAGCATCACAGCTGGCGCAA 2249  
Qy 721 GlnLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740  
Db 2250 GAGCTGTTTTTGAATTAAGATACAGCCAGGCTGATGCCCTGAGTATGTGGCATCGAA 2309

Qy 741 ArgGluMetGluIlePro 746  
Db 2310 AGAGAAATGGAATTCCT 2327  
RESULT 3  
AR473734  
LOCUS AR473734 2600 bp DNA linear PAT 20-FEB-2004  
DEFINITION Sequence 1 from patent US 6689583.  
ACCESSION AR473734  
VERSION AR473734.1 GI:42712196  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1. (bases 1 to 2600)  
AUTHORS Jenuwein,T., Laible,G., O'Carroll,D., Eisenhaber,F. and Rea,S.  
TITLE Chromatin regulator genes  
JOURNAL Patent: US 6689583-A 1 10-FEB-2004;  
Boehringer Ingelheim International GmbH;  
DEX;  
FEATURES  
location/Qualifiers  
source 1..2600  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0  
US-10-773-302-2 (1-746) x AR473734 (1-2600)  
Qy 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
Db 90 ATGGGCCAGACTGGGAAGAAATCTGAGAGAGAGACCACTTGTGTGGAGAGCGGTGAAA 149  
Qy 21 SerGluThrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
Db 150 TCAGAGTACAGACCACTGACAGACAGCTCAAGAGGTTCAGACAGAGCTGATGAAGTAAAG 209  
Qy 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnLysTrp 60  
Db 210 ATGTTTGAATTCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAACCAAGANTGG 269  
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 270 AAACAGGAAAGATACAGAGCTGTGCACATCTGTGAGCTCATTTGCCGGAGCT 329  
Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 330 AGGAGGTGTTGGGACCAAGTGAATTTTCAACACAGATCATCTTAAGACT 389  
Qy 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
Db 390 CTGAATGAGTGTCTTCAGTACCAATATGATTTCTTGTTCCCTCAAGCAGCAATTTT 449  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
Db 450 ATGGTGAAGATGAACGTTTTCATTAACAATTCCTTAATGAGGAGATGAAGTTTATGAT 509  
Qy 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
Db 510 CAGATGTACTCTTCATTGAAGAACTAATAAAAAATTAATGATGAGAAAGTACAGGGGAT 569  
Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 570 AGAAGATGGGTTTATTAATGATGAAGAAATTTTGTGAGTTGGTGAAGTGCCTTGGTCAA 629

OY		181	TyzAmnAphrApArPhrApArPhrApArgIyAphArPrOgIuGIuArGluGIuLySglN	200
Dd		630	TATATAATGTAGTGAAGATGATGATATGATGAGAAGCTCCTGAAGAAAGAAAGAAAGACG	689
OY		201	LysAPRleuGIuASrhNIeArGrAPhryLysGIuSerARProArGLyPheProser	220
Dd		690	AAGATCTGGAGGATCCCGAGATATTAAGAAACCGGCCACTTCGGAAATTCTCTCT	749
OY		221	AspLySIephelIuAlIeserSerMetPheProAspLySglYThrAlagIuGIuLeu	240
Dd		750	GATTAATAATTTTGAAAGCATTTTCCTCAATGTTTCCAGATTAAGGCACAGCAAGAAACTA	809
OY		241	LysgIuLyTYrLYegIuLeuthrGIuIngInLeuPrOGIyAlAleuProProGIuCyS	260
Dd		810	AAGAAAAATATTAAGAACTCACCCAGACAGAGCTCCAGGGGCACTTCCTCTGAAGT	869
OY		261	ThrProAmIIeaSrGIyProsrnAlIysSerValGIuArGluGIuInSerIenHIsSer	280
Dd		870	ACCCCAACATAGAGAACCAATCTAAATGTGTCAAGAGAACAAAGCTTACATCC	929
OY		281	PheNIsthrLeuPheCYarGrArGYsPheLySTyrAsrCYsPheLeuHIspRoHeNI's	300
Dd		990	TTTTATAGGCTTTTCTGTAGGGAGTGTTTAAATATGATGCTGTTCTCAATCCTTTTCAAT	989
OY		301	AlathrProArnthTYrLYsarGLySAanthrgIuthralAleuAsrAnlysrCYoS	320
Dd		990	GCAACACCACACTTATATAGCGGAACAACAGAAACAGCTCTGACAAACAACCTGT	1049
OY		321	GIyPrOGInCYeTYrGIuHIleuNGluGIyAlalySGIuPheAlAlAleuthrAla	340
Dd		1050	GGACCAAGTGTTACAGATTTTGAAGGAGAGAAAGATTTGGCTGTGACTTCACCGCT	1109
OY		341	GIuArGIleLythrProProLyArGrProGIyGIyArGrArGrGIyArGleuProAsn	360
Dd		1110	GAGCGGATPAAGACCCCAACAAAAGTCCAGAGGCCCCCAAGAAAGACGGCTCCCAAT	1169
OY		361	AmsrSerArGrProSerThrProThrIIeAnValIeuGIuSerLYasrPthrAsrSer	380
Dd		1170	AACATATCACGGCCACGACCCCACTTAATGTGCTGAATCAAAGATACAGACAGT	1229
OY		381	AsrArGrGIuAlagIYThrGIuThrGIyGIuAnAsrAPhySGIuGIuGIuGIuLyS	400
Dd		1230	GATAGGGAAGCAGGAGCTGAACCGGGGAGAGAAATGTATTAAGAAAGAAAGAAAGAG	1289
OY		401	LysAPRGIuThrSerSerSerSerGIuAlIAnsSerArGYSGInThrProIIelySet	420
Dd		1290	AAAGATGAACCTTCAGACTCCTCTGAAGCAANTTCCGATGCANAKCCCATTAAGATG	1349
OY		421	LysProAmIIeGIuProProGIuAnValGIuTPSerGIyAlagIuAlAsrMetPhe	440
Dd		1350	AAGCCAATATATGAACCTCTGAGAGATGTGAGTGGTGGTGAAGCCTCAATGTTT	1409
OY		441	ArgValIeuIIeGIYThrTYrTYrArAsrPheCYeAlAlAleArGLeuIIeGIYThr	460
Dd		1410	AGAGCTCCATTTGGCACTTACTATGACAATTTCTGTGCAATTTGATGATTAATGGACC	1469
OY		461	LYeThrCYsArGrGIuValTYrGIuPheArGVallySGIuSerSerIIeAlAsrAla	480
Dd		1470	AAAAATGTAGACAGGTATGATGATTTAGAGTCANAAGATCTAGATCATAGCTCCAGCT	1529
OY		481	ProAlagIuAsrValAsrPthrProBrArGLyLySlySArGLyVHIeArGLeuITrAla	500
Dd		1530	CCCCGTGAGATGTGATATCTCTCCAGAGAAAAAGAGAGAAACAACGGGTGGGGCT	1589
OY		501	AlAhISCYeArGLyIIeGIuLeuLylyAsrGIySerSerAsrHIVAlTYrAsrTYr	520
Dd		1590	GCACACTGCAGAAACATACAGTGAAGAAAGACCGGCTCTCTAACCAAGTTTACACAT	1649
OY		521	GIuPrOCYSAPRHIEPRoArGLINPrOCYSAPRSerSerCYsPrOCyValIIeAlagIn	540
Dd		1650	CACCCCTGTGATCATCCACAGGACGGCTTGTGACAGTTGTGCCCCTTGTGTATAGACAA	1709
OY		541	AsnPheCYgIuLySPheCYeGIuCYsSerSerGrGIuCYgInAsrAtyPheProGIyCyS	560

[illegible]

Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0

US-10-773-302-2 (1-746) x AK333675 (1-2600)

Qy 1 MetGlyInThrGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
Db ATGGGCGGACGCTGGGAAAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAACGGTAAAA 149  
Qy 21 SerGlyTrpMetArgLysArgLysLysArgPheArgArgLysArgLysLysSer 40  
Db TCAGAGTACATGACCACTGAGCAGCTCAAGAGTTTCAAGACAGAGCTGATGAAGTAAAGGT 209  
Qy 41 MetPheSerSerAsnArgLysLysLysLysLysLysLysLysLysLysLysLysLys 60  
Db ATGTTTACGTTCCATTCGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATGG 269  
Qy 61 LysGlnArgArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80  
Db AAACGCGAAGATACAGCTGCTGCACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 329  
Qy 81 ArgGlyCysSerValThrSerAspLysAspPheProThrGlnValLysLysLysLys 100  
Db AGGAGTGTTCGGTGAACGAGTGAATTTTCCAAACAAGTATCCCATTAAGACT 389  
Qy 101 LeuAsnAlaValAlaSerValProLysMetTyrSerTrpSerProLeuGlnLysAsp 120  
Db CTGAATGAGTGTCTCTGATGACCAATATGATTTCTTGTCTCTCTCTCTCTCTCTCTCT 449  
Qy 121 MetValGluAspGlyThrValLysLysLysLysLysLysLysLysLysLysLysLys 140  
Db ATGGGGAAGATGAAAGCTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTACAT 509  
Qy 141 GlnAspGlyThrPheLysLysLysLysLysLysLysLysLysLysLysLysLysLys 160  
Db CAGATGTGACTTTCATTTGAAGAACTAATTAATAATTAATTAATTAATTAATTAATTAAT 569  
Qy 161 ArgGlyCysGlyPheLysAsnAspGlyLysPheValGlnLysLysLysLysLysLys 180  
Db AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGTGTGAATGCCCTTGTGTGA 629  
Qy 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
Db TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
Qy 201 LysAspLeuGlnAspPheLysArgAspAspLysGlyLysArgProProArgLysPhePro 220  
Db AAAGATCTGAGAGATCACCGAGATATTAAGAAACCGCCGACCTCGGAAATTTCTCTCT 749  
Qy 221 AspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 240  
Db GATTAATATTTTGAAGCAATTTCTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 809  
Qy 241 LysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 260  
Db AAGGAAATATTAAGAACTCAACGAAACAGCTCCCAAGGCGCACTTCTCTCAAGATGT 869  
Qy 261 ThrProAsnLysAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLysLys 280  
Db ACCCCCAACATAGATGACCAATCTAATCTGTTCAGAGAGACCAAAAGCTTAACATCC 929  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLysLysLysLys 300  
Db TTTCTATAGCTTTTCTGTAAGGAGATGTTTAAATATGATGCTGTTCTTCAATCTTTTCA 989  
Qy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
Db GCAACACCAACACTTATTAAGGGAAGAAACAAGAAACAGCTCTAGACAAACAACTTGT 1049  
Qy 321 GlyProGlnCysTyrGlnHisLysLysLysLysLysLysLysLysLysLysLysLys 340  
Db GACCAACAGTGTACACATTTTGAAGGAGCAAAAGAGTTTGTGCTGCTCTCAACGCT 1109

Qy 341 GlnArgLysLysThrProProLysArgProGlyLysLysArgArgLysLysLysLys 360  
Db GAGCGGATTAAGACCCCAACAAACGTCAGAGAGCGCGCAAGAGAGAGCGGCTTCCCAAT 1169  
Qy 361 AsnSerSerArgProSerThrProThrLysAsnValLysLysLysLysLysLysLysLys 380  
Db AACATGACAGGCGCCACACACCCCAACATTAATGCTGGAATCAAAAGATACAGACAGT 1229  
Qy 381 AspArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 400  
Db GATGAGAACAGGAGATGAAACGGGGGAGAGACATGATTAAGAAAGAAAGACAGAG 1289  
Qy 401 LysAspGlyThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 420  
Db AAAGATGAACCTTCGACCTCTCTGAAAGCAATTTCTGCTGTCTCAACCAATTAACATG 1349  
Qy 421 LysProAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 440  
Db AAGCAATATTAAGACCTCTGAGAAATGAGAGTGAAGTGTGCTGAAGCTCAATGTTT 1409  
Qy 441 ArgValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 460  
Db AGAGTCTCTATGAGCACTTATGATGACATTTCTGTCCATTTGCTATGTTAATGGAGCC 1469  
Qy 461 LysThrCysArgGlnValTyrGlnPheArgValLysLysLysLysLysLysLysLys 480  
Db AAAACATGATGACAGGTATGAGTTAAGTCAAAAGATCAATGATCATATGCTCCAGCT 1529  
Qy 481 ProAlaGluAspValAspThrProProArgLysLysLysLysLysLysLysLysLysLys 500  
Db CCCGCTAGATGAGTGTGATCTCTCAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1589  
Qy 501 AlaHisCysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 520  
Db GCACATGCGAAGAAAGATACAGCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1649  
Qy 521 GlnProCysAspHisAspArgGlnProProCysAspSerSerCysProCysValLysLys 540  
Db CAACCTGTATGATTCACACCGGACCTTGTGACAGTTCGCTGTGTGTGTGTGTGTGTGT 1709  
Qy 541 AsnPheCysGlyLysPheCysGlnCysSerSerSerSerSerSerSerSerSerSerSer 560  
Db AATTTTGTGAAAGTTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1769  
Qy 561 ArgCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 580  
Db CGCTGCAAGGACAGTGCACCAACCAAGCAAGGCGCTGCTGCTGCTGCTGCTGCTGCTG 1829  
Qy 581 AspProAspLeuCysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 600  
Db GACCTGACCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1889  
Qy 601 CysLysAsnCysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 620  
Db TGCAGAACTGCAATATTCAGGGGGCTCCAAAAGCAATATTCCTGCGACCAATCTGAC 1949  
Qy 621 ValAlaGlyTrpGlyLysPheLysLysLysLysLysLysLysLysLysLysLysLys 640  
Db GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGCAAGAAATGAATCATCTCAGAA 2009  
Qy 641 TyrCysGlyGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 660  
Db TACTGTGAGAGATTAATTTCTCAAGATGAGCTCAAGAAAGGGAAGGGAATGATGAATA 2069  
Qy 661 TyrMetCysSerPheLysPheAsnLysAsnAspPheValValAspAlaThrArgLys 680  
Db TACATGTGCAAGCTTCTGTTCACTTAACATATATTTTGTGGAGTGAACCCGAGAG 2129  
Qy 681 GlnAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 700  
Db GGTAAACAAATTCCTTTTGGAAATCATTCGTTAATCCAAATCTGTATGCAAAAGTTATG 2189

**QY** 701 MetValaenGlyAspHisArgIleGlyLeuPheAlaValArgAlaIleGlnThrGlyGlu 720  
**Db** 2190 ATGCTTAACGGTGAATCAAGATAGTATTTTGGCAAGAGCCATCCAGCTGGCCAA 2249  
**QY** 721 GlnLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740  
**Db** 2250 GAGCTGTTTTTGGATTACAGATACAGCCAGCTGATGCCCTGAAGATATGTCGACATCGAA 2309  
**QY** 741 ArgGluMetGluIlePro 746  
**Db** 2310 AGAGAAATGGAATCCCT 2327  
**RESULT 5**  
**LOCUS** HSU61145 2600 bp mRNA linear PRI 11-UTL-1997  
**DEFINITION** Human enhancer of zeste homolog 2 (EZH2) mRNA, complete cds.  
**ACCESSION** U61145  
**VERSION** U61145.1 GI:1575348  
**KEYWORDS**  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens (human)  
**REFERENCE** 1 (bases 1 to 2600)  
**AUTHORS** Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Leberer, A., Popkin, D., Pillus, L. and Jenuwein, T.  
**TITLE** Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres  
**JOURNAL** EMBO J. 16 (11), 3219-3232 (1997)  
**PUBMED** 9214638  
**REFERENCE** 2 (bases 1 to 2600)  
**AUTHORS** Laible, G., Leberer, A. and Jenuwein, T.  
**TITLE** Direct Subunit  
**JOURNAL** Submitted (18-JUN-1996) I.M.P., Dr. Bohrgasse 7, Vienna A-1030, Austria  
**FEATURES**  
**source** Location/Qualifiers  
 1..2600  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_line="Bm-B"  
 /cell\_type="B-cell"  
 1..2600  
 /gene="EZH2"  
 /note="EZH2"  
 /note="Allele: EZH1; homolog of the Drosophila polycomb group protein enhancer of zeste encoded by GenBank accession Number U00180"  
 /codon\_start=1  
 /product="enhancer of zeste homolog 2"  
 /protein\_id="AAC51520.1"  
 /db\_xref="GI:1575349"  
 /translation="MGQTGKSEKPVCKRKRYKSEYKRLQKRPRADEKMFSSNRQKLEKTEILNQEKQRIQPHILTSVSLRTRCSVSDLDPTQVPLKTLNAAVAVPIMYSPLQONFMVDETVLHNI PYMGDEVLDQDFIBELKNYGVKAGDRECFINDEIFELVNALGOYDDDDDDGDPBEREKQKLEHDKSRPRPKFPSDKIFBAISSMFPDKTAELKEKYKELTEQALPGALPCEPIIDGNPAKSVQREO SLHSFHTLCRCRCFVDCGTLHFHATPNTYRKNTETALDNKPGQCYOHLBAKKEF AALTAERIKTPPRKPGRRRLPNNSRSTPTINTLESCTDSDBAERTGGENNDKEERKDETSSESSKANSRCQTPIKKPNLEPENVBWSGASMFRLIGTYDNFCALARLIGTKRCQVYEFKVSIIAPAPAEVDTPRKQRKRLMAACRIQLKKDSNNVNYVPCDHPKDCSCPCVIAQNFCEKQSCSECCNFPKPCRAQCNKQPCYLAABEDPDLCTCGADHDSKNSCKNSIQSGSKHLLAPSDVAGNGIFIDPVOKNEFISEYGGRIISODEADRGKVDKWCSPFLNNNDPVADATKKNKIRFANHSVNPCTAKVMVNGDHRIGIPAKAIIQTGSELFPDITRYSQADALKYGISREMRIP"  
**ORIGIN**  
**Alignment Scores:**  
**Pred. No.:** 0 Length: 2600

**Score:** 4051.00 **Matches:** 746  
**Percent Similarity:** 100.0% **Conservative:** 0  
**Best Local Similarity:** 100.0% **Mismatches:** 0  
**Query Match:** 100.0% **Indels:** 0  
**Db:** 5 **Gaps:** 0  
**US-10-773-302-2 (1-746) x HSU61145 (1-2600)**  
**QY** 1 MetGlyGlnThrGlyValLeuSerGlnLysGlyProValCysTRPArgLysArgValLys 20  
**Db** 90 ATGGGCCAGACTGGGAAAGAAATCTGAGAAAGGACCAAGTTGTGGCGGAAGCGTAA 149  
**QY** 21 SerGluTyrMetCysArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
**Db** 150 TCAGAGTACATCCGATGAGTACAGTCAAGAGTTCAACACAGCTGATGAAGTAAAGGT 209  
**QY** 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnLys 60  
**Db** 210 ATGTTAGTCCAACTGTCAGAAATTTGGAAAGAACGAAATCTTAAACCAAGATCG 269  
**QY** 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerLeuAspGlyThr 80  
**Db** 270 AAACAGCAGAGATACACCTGTGCACATCTGACTTGTGAGCTCATGTGCGGAGCT 329  
**QY** 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
**Db** 330 AGGAGTGTTCGGTACCAAGTACATGATTTTCCAAACCAAGTATCCCATTAAGACT 389  
**QY** 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnLysAsn 120  
**Db** 390 CTGAATGAGTGTCTTCAATGATCCATTAATGATTTCTGTCTCCCTACAGCAGATTTT 449  
**QY** 121 MetValGluAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
**Db** 450 ATGTGGAAAGTGAACAGTTTACATTAACATCTCTTAATAGGAGATGAATTTTAGT 509  
**QY** 141 GlnAspGlyThrPheIleGlnLysLeuLysAsnTyrAspGlyLysValHisGlyAsp 160  
**Db** 510 CAGATGATCTTTCATTTAGTAAGACTAATTAATAATTTATGATGGAAAGTACACGGGAT 569  
**QY** 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGln 180  
**Db** 570 AGAGATGTGGGTTATTAATGATGAATTTTGTGGAGTGTGTAATGCCCTTGCTCA 629  
**QY** 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluLysGln 200  
**Db** 630 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
**QY** 201 LysAspLeuGluAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
**Db** 690 AAAGTCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
**QY** 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluLys 240  
**Db** 750 GATAAATTTTGAAGCATTTTCCATATGTTTCCAGATTAAGGCGACAGCAAGAACTA 809  
**QY** 241 LysGluLysTyrLysGlnLeuThrGlnGlnLysProGluValAlaLeuProProGlu 260  
**Db** 810 AAGGAAAAATVTAAGAACTACACCGAACAGAGCTCCAGGCGCACTTCTCTTAAGT 869  
**QY** 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
**Db** 870 ACCCCACATAGATGACCAAGGCTAAATCTGTTCAGAGACCAAGGCTTACATCC 929  
**QY** 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisPheProHis 300  
**Db** 930 TTTGATAGCTTTTCTGAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 989  
**QY** 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
**Db** 990 GCAACACCCCAACACTTATATAGCGAAGAAACAGAAACAGCTTATAGAACAAACCTTGT 1049  
**QY** 321 GlyProGlnCysTyrGlnHisLeuGluLysValLysGluPheAlaIleAlaLeuThrAla 340

Db	1050	GGACCAAGTGTACCAAGATTGGAGGAGCAGAGAGTTTGCTGCTTCACCGCT	1109
Qy	341	GIuArGIleYrThrPrOProLyArPProGIyArGArGArGIyArGIeUProAr	360
Db	1110	GAGGGATTAAGACCCCAACAAACGTCCAGGAGCGCAGAGAGCAGCGCTCCAA	1169
Qy	361	AAnSerSerArPProSerThrProThrIleAenValleuGIuSerLyArPThrApsSer	380
Db	1170	AACGTAGCAGAGCCCAAGACCCCACTTAATGTGCTGGAAATCAAAAGATACAGACGT	1229
Qy	381	ASpArGIuAlaGIyThrGIuThrGIyGIyGIuAenAaAPLyGIuGIuGIuLyS	400
Db	1230	GATGGGAGAGCAGGACTGAACCGGGGAGAGAAACAATGATTAAGAAAGAGAGAG	1288
Qy	401	LYSAAPGIuThrSerSerSerSerGIuAlaAAnSerArGYSGInThrProIleYsMet	420
Db	1290	AAAGATGAACCTTGAGAGCTCCTCTGAAACAATCTCGGTGTCAAAACCAATAAAGTG	1349
Qy	421	LYSProAenIleGIuProProGIuAenValGIuTPSerGIyAlaGIuAlaSerMetPhe	440
Db	1350	AAGCCAAATATTTGAACCTCTGAGAAATGTGAGTGTGAGTGTGCGAAGCCCTCAATGTTT	1409
Qy	441	ArgValleuIleGIyThrTrpThrAAnPheCySAIleAlaArgleuIleGIyThr	460
Db	1410	AGAGCTCCATTTGGCACTTACTATGACAAATTTCTGTGCATTTGCTAGGTTAATGGAGCC	1465
Qy	461	LYSThrCYSArGIuValIyTrGIuPheArgValIySGIuSerSerIleAlaAProAla	480
Db	1470	AAAACATTAAGACAGGTGATGAGTTTGAAGTCAAAAGATCTAGCATCATAGCTCAGCT	1525
Qy	481	ProAlaGIuAPValAAPThrProProArgLySValySAlaArgLeuThrAla	500
Db	1530	CCCCGTGAGATGTGGATATCTCTTCAAGGAAAAAGAGAGAAACACCGTGTGGGCT	1585
Qy	501	AlaHISCYSArGIySIIeGIuLeuYelYSAAPGIySerSerAAnHISValIyThrAnIyTr	520
Db	1590	GCACCTGCAGAAAATACAGCTGAAAAAGACGGCTCTCTTACCAAGTTTACAACTAT	1649
Qy	521	GIuProCYSAAPHISProArgGIuProCYSAAPSerSerCYSProCYSAValIleAlaGIu	540
Db	1650	CAACCTGTGATCATTCACGAGCGCTGTGACATTTGTGCCCTGTGTGTATACAGACA	1709
Qy	541	AAnPheCYSGIuLyPheCYSGInCYSSerSerGIuCYSGInAAnArgPheProGIuCYs	560
Db	1710	AATTTTGTGAAAAGTTTGTTCAAATGTAGTTCAGAGTGCAGAAACCGCTTCCGGGAGTGC	1765
Qy	561	ArgCYSAlySAlaGIuCYAAnThrIlySGInCYSProCYTrpLeuAlaValArgGIuCYs	580
Db	1770	CGCTGCAAAGCACAATGTCAACACCAAGACGTCGCCGTCTGCTGCGAGAGGT	1825
Qy	581	AAPProAPLeuCYeLeuThrCYSGIyAlaAlaAePHISTrpAPSerSerLySAenValSer	600
Db	1830	GACCTTGACCTGTGTCTTACTTTGTGTGAGCGCTGTACATTTGGAGACAGTAAATGTGTCC	1885
Qy	601	CySlySAenCYSerIleGIuArgGIySerIySlySHIleuLeuLeuAlaProSerArP	620
Db	1890	TGCAGGAAGTGCAGATTCAGGGGGCTCCAAAAGACATCTATGCTGGCACCATCTGAC	1945
Qy	621	ValAlaGIyTrpGIyIlePheIleYSAAPProValGIuLySAAnGIuPheIleSerGIu	640
Db	1950	GTCGAGAGCTGGGGATTTTATCAAAAGTCTGTGAGAAAAGAAATTCATTCAGAA	2005
Qy	641	TyrCYSGIyGIuIleIleSerGIuAAPGIuAlaAAPArGArGIyLySValIyZArAPLyS	660
Db	2010	TACTGTGAGAGATTAATTTCTCAATGAGAGCTGCAGAGAGGAGAAAGTGTATATATA	2065
Qy	661	TyMetCYSerPheLeuPheAAnLeuAAnAAnAPheValAAlaPAlaThrArgLyS	680
Db	2070	TACATGTGCAAGTTTCTGTTCAACTTGAACATGATTTTGTGTGTGAGACCAACCGCAAG	2125
Qy	681	GIyAenLySIIeArgPheAlaAAnHISSerValAAnProAenCYTrpAlaLySValMet	700

Dn		2130	GGTAAACAATAATGCTTTGGAAATCATTCGCGTAATCAACACTGCCTATGCCAAAAGTATG	2189
Oy		701	MetValJansglYasphISatgllleglylIephelalYahtayghlaaleglnthnglyglu	720
Dn		2190	ATGGTTAACCGTGATCACAGAGTAGGTATTATTGGCCAAAGAGCATTCCAGACTGGCGAA	2249
Oy		721	GlueuaphepheApPyrArGrTYrSerglnAlAsphAlalauleuSYrVaAlglYlIeglU	740
Dn		2250	GAGCTGTGTTTTTATTTCACAGTACAGCCAGGCTGATCCCTGAAGTATGTCCGATCGAA	2309
Oy		741	ArgGlumetGluilePro	746
Dn		2310	AGAGAATGGAAATCCCT	2327
RESULT 6				
LOCUS		ABI179385		
DEFINITION		Macaca fascicularis testis cDNA clone: Qtea-19821, similar to human enhancer of zeste homolog 2 (Drosophila) (EZH2), transcript variant		
ACCESSION		ABI179385		
VERSION		ABI179385.1		
KEYWORDS		oligo capping; fis (full insert sequence)		
SOURCE		Macaca fascicularis (crab-eating macaque)		
ORGANISM		Macaca fascicularis		
REFERENCE		Bukacinski, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.		
AUTHORS		Osada,N., Hirata,M., Tanuma,R., Kusuda,J., Hida,M., Suzuki,Y., Sugano,S., Gojobori,T., Shen,C.-K.J., Wu,C.I. and Hashimoto,K.		
TITLE		Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis Between Human and Cynomolgus Monkey cDNAs Mol. Biol. Evol. 22 (10), 1976-1982 (2005)		
JOURNAL		15944441		
PUBMED		2		
REFERENCES		International consortium for macaque cDNA sequencing and analysis DNA sequences of macaque genes expressed in brain or testis and its evolutionary implications Unpublished		
CONSTRM		3 (bases 1 to 2619)		
AUTHORS		Hashimoto,K., Kusuda,J. and Sugano,S.		
TITLE		Direct Submission		
JOURNAL		Submitted (10-MAY-2004) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shijuku-Ku, Tokyo, 162-8640, Japan		
COMMENT		(E-mail:hashimkuh.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)		
		The International consortium for macaque cDNA sequencing and analysis consists of:Department of Virology and Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; Division of Genetic Resources, National Institute of Infectious Diseases of Japan, Tokyo, Japan; National Health Research Institute, Taipei, Taiwan; Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan; Department of Ecology & Evolution, University of Chicago, Chicago, IL, USA, Center for Information Biology, National Institute of Genetics of Japan, Mishima, Japan. Clone distribution: clone distribution information can be found at: http://www.nih.go.jp/yoken/genebank/		
		Lab host: TOP10		
		Vector: pME18S-F13 (ACC.No. AB009864)		
		R. Site1: DraIII (CACTGTGAG)		
		R. Site2: DraIII (CACCATGGG)		
		Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTCCTCTTTTATTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-F13. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries		



were constructed by oligo-capping method. Libraries were made from:

Qc6: cerebellum cortex  
Qnpa: parietal lobe  
Qtra: temporal lobe right  
Qfla: frontal lobe left  
Qmoa: medulla oblongata  
Qbsa: brain stem  
Qora: occipital lobe right  
Qesa: testis

Custom primers were used for 5' and 3'-end sequencing. The full-insert sequencing was done by primer-walking method using ABI DNA sequencer.

# FEATURES

## source

## CDS

Location/Qualifiers  
1..2619  
/organism="Macaca fascicularis"  
/mol\_type="mRNA"  
/db\_xref="taxon:9541"  
/clone="Qc6a-18821"  
/sex="male"  
/clone\_lib="macaque cDNA library Qc6a"  
/dev\_stage="adult"  
93..2333  
/note="unnamed protein product; Homo sapiens enhancer of zeste homolog 2 (Prosophiila) (EZH2), transcript variant 2, mRNA, RefSeq: NM\_152998.1"  
/codon\_start=1  
/protein\_id="BAB02436.1"  
/db\_xref="GI:67972188"  
/translation="MGQTGKSEKGPVCRKRVKSEYMRRLQKRPBRADDEVKSWSSNRKILBERTILNDEKORRIQPIHLISVSLRGTRCSVSDLPPOVYPLKTLNAYASPIVSWSPQLQONFMVDEDTLNIPIWGDRLDDGDFRTEILIKNYDGKYNHDRECGINDEIFVELYNALQYVNDDDDDGDDPERREKQKLEHRDDKSRPRKRPDSKIFEAISHPDKGTABELKEKYLEQQLFGALPBECTPIIDGNASVQREO SLASHPTLFCRCRCPYDCFLPHFATPNAKNTETALDNPCOPQCYOHLGAEKFAALTAERIITKTPPKRPGRRRLPNNSRPSTPTINYLESDTSDSEAGETGGENNDKEBKEDETSSSSBANSRCOTPIKKPNLEPENWESGASMSRVLIGTYDNFCATARLIGTKCROVPEPRVSESILAPAEVDTPRKRKHGLMAHCRKIQIKKDGSSNHYNTQPCDHPROPDSSCPVTIAGNFCRKCQCSSECCNRPFGRCRAQCNTQPCYLAVERCDPLCTCGADHDSKNVSKNSIQHGSKHLLAPSDVAGNFIKDPVKQKEFISYCGEILISQDEARGRKYDKRCSPLFNLDNPFVDDATKGNKIRFANHSVNPFCYAKVMVWNGDRIGIFAKRAIQGBELFDYRYSQADMLKTVGIREMEIRP"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 2619  
Score: 4046.00 Matches: 745  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.9% Indels: 0  
DB: 5 Gaps: 0

US-10-773-302-2 (1-746) x ABI179385 (1-2619)

QY 1 MetGlyGlnThnGlyLysSerGlyLysGlyProValCysTTPArgLysArgValLys 20  
DB 93 ATGGGCGACGACTGGGAGAAATCTGAGAGGACCAATTGTTGGCGGAGGTGTAATAA 152  
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
DB 153 TCGAGTACATCCGACTGACGACCTCAAGAGGTTCAACGAGCTGATGATTAAGAGCT 212  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThnGluIleLeuAsnGlnLutTrp 60  
DB 213 ATGTTTAGTTCCAAATCGTCAGAAATTTTGGAAAGACGGAATCTTAAACCAAGATCG 272  
QY 61 LysGlnArgArgGlnIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 273 AAACAGCGAAGATGACGCTGTGACATCTTCACTTCTGTAGCTCATTTGGCGGAGCT 332  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThnGlnValIleProLeuLysThr 100

DB 333 AGGAGTGTGGGTGACCAAGTACCTTGAGTTTCCACACAAATGATCCATTAAAGACT 392  
QY LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnIleAsnPhe 120  
DB 393 CTGAATGACGTTGCTTCAGTACCACTAATGATTTCTGGTCTCCCTTCAGCAAAATTT 452  
QY MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 453 ATGCTGGAAGATGAAACGTTTTCATACATTCCTCGTATATGAGGAGATGAAGTTTAGAC 512  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 513 CAGATGTGACTTTCATTGAAGAACTAATTAATAATTTATGATGAGAAAGTACACCGGAGAT 572  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLysValAlaAsnAlaLeuGluYgln 180  
DB 573 AGAGATGTGGGTATTAATGAATGAATTTTGTGGAGTTGCTCAATGCTTGTGCTCA 632  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
DB 633 TTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 693 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 752  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 753 GATTAATAATTTTGAAGCAATTTCTCCAGTGTTCAGATTAAGGGGACAGCAAGAAACTGA 812  
QY 241 LysGluLysTyrLysGluLeuThnGluGlnGlnLeuProGlyAlaLeuProProGluCys 260  
DB 813 AAGGAAAAATTTAAGAACTACACGACGACGACGCTCCAGGCGACTTCTCTGTAATGT 872  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerIleHisSer 280  
DB 873 ACCCCCAATATAGTAGACCAAAACGCTAAATCTGTTGAGAGAGCAAAAGCTTACATCC 932  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 933 TTTCAATAGCTTTCTGTAGGCGATGTTTAAATATGATCTGTTCAATCTTCAATCTTTTCAAT 992  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThnGluThrAlaLeuAspAsnLysProCys 320  
DB 993 GCAACACCAACGCTTATTAAGCGAGAAACACAGAAACAGCTTTGACACAAACCTTGT 1052  
QY 321 GlyProGlnCysTyrGlnHisLeuGluGluValAlaLysGluPheAlaAlaAlaLeuThrAla 340  
DB 1053 GGACCAAGTGTACACGACTTTGGAGGAGCAAAAGAGTTCCGTCGTCTCTCACTGCT 1112  
QY 341 GlnArgGlyLeuThrProProLysArgProGlyGlyArgArgGlyArgLeuProAsn 360  
DB 1113 GAGCGGATCAAGAACCCCAACAAACGCCAGAGAGCCGACAGAGGAGCGGCTTCCCAAT 1172  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleGluSerLysAspThrAspSer 380  
DB 1173 AACGTAACAGGCCACGACACCCCACTTAATGTGCTGGAATCAAGAGATACAGACAGCT 1232  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluLys 400  
DB 1233 GATGAGGAAGACGAGATGAAACGGGGGAGAGAAACATGATTAAGAAAGAAAGAGAG 1292  
QY 401 LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1293 AAAGCGAAACCTTGAGCTCTCTGAGCAAAATCTCCGTGTCAAAACCAATTAAGATG 1352  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1353 AAGCAATATATTAACCTCTGAGATGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1412  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1413 AGAGTCTCATTTGGACATTATGATGACAAATTTCTGTGCAATTTGCTAGGTTAATGGAGCC 1472



Qy	461	LYSTHRCYBAAGGGINVALTYRGILUPHEARGVALLYSGIUSERSERIIEIIEAIPROA	480
Db	1473	AAAACTGTACACAGTGTATGAGTTTAAAGTCAAAAGATTCATGATCATGCTCCAGCT	1532
Qy	481	PROAAGIUAAPVALAAPTTHRPROPROARGLYSLYSLYSAARGLYSHIISARGLEUTPALA	500
Db	1533	CCTGCTGAGGATGTGTGACATCTCTCCAGAGAAAAAGAAAGAGAAACCCGTTGTGGCT	1592
Qy	501	ALAHISCYBARGLYSEIIEGLINLEULYSLYSAAPGLYSERSEARSHISVALITYRANTYR	520
Db	1593	GCAACACTGCAGAAAGATACACTGCAGAAAAAGACCGCTCTTAACCAATGTTTACAACTAT	1652
Qy	521	GLINPROCYBARPHIISPROARGLINPROCYBARPHISERISERYSPROCYVALIIEALAGLIN	540
Db	1653	CAACCCGTGTGATCATCCACGGAGCGTTGTACAGTTGTGTCCCTTGTGTATGACAA	1712
Qy	541	AAPNHCYSGIULYSPHPCYSGELINCYSSERISERYSGINLAEMARGPHEPROGLCY	560
Db	1713	AATTTTGTGAAAAAGTTTGTGCATGTGATTCAGAGTGTCAAAACCGCTTCCGGAGAGC	1772
Qy	561	ARGCYSLYVALAGLINCYBARNTHIRYSGINCYSPROCYTYTLEUALIIVALARGGLI	580
Db	1773	CGCTGCAGAACGACAGTGCACACCAACGACAGGCCGCTCTACCTGCGCGTCCGAGAGGT	1832
Qy	581	AAPPROABAPLEUCYSLLEUTHRCYSGIYVALIIEAAPHIISTRAPASERLYSAPNVALSER	600
Db	1833	GACCTTGACCTGTGTCTACTGTGTGGACCCGTGACCATGTGGACAGTAAAAAGTGTCC	1892
Qy	601	CYSLYSAANCYSSERIEGLINARGGLYSEIRYSLYSHIISLEULEUALIIPROSERASP	620
Db	1893	TGCAAGAACTGACAGTATTCAGCGGGGCTCCAAAAAGCATCTATTGCTGGACCATCTGAT	1952
Qy	621	VALIIEAGLYTTPGLYIIEPHEIIELYSAPPROVALGINLYSANGIUPHEIIESERCIU	640
Db	1953	GTGGCAGGCTGGGGATTTTATCAAAATCCTGTGCGAAAAATGAATTCATCTCAAAA	2012
Qy	641	TYTCSGGLYGINIIEIIESERCIINAPSGIULIIEAAPARGARGGLYLYSVALITYRASPLYS	660
Db	2013	TACTGTGAGAGATTAATTTCTCAGATTAAGCTGCACAGAGGGAAAGTATATGATTA	2072
Qy	661	TYMETCYSSERPHLEUPHEASNULEUANSHAPHEVALIIEASPALIITHRARGLYS	680
Db	2073	TACATGTGCAGCTTCTGTCTCAACTGAAACATGATTTTGTGGTGAATGCAACCGCAG	2132
Qy	681	GLYASNLYSIIIEARGPHEALIAASHISSEERVALIIEANPROANCYRTYRIILYLYSVALMET	700
Db	2133	GGTAAACAAATTCGTTTGGCAATCATTCGGTAAATCCAAATCTGTATGACAAAAGTTATG	2192
Qy	701	METVALAENGILYASPHIISARGIIEGLYIIEPHEALIALYSARGALAIIEGLINTHRCIYGLU	720
Db	2193	ATGTGTAAACGGCGATCACAGATAGATGATTTTGTCCAGAGAGACCATCACACATCGTGA	2252
Qy	721	GIULEUPHEPHEASPTYRARGTYRISERCINIAAEPALAEULYRTYRVALIGLYIIEGLU	740
Db	2253	GAGCTGTTTTTGTGATTACAGATACAGCAGCGGTGATGCCGTGAAGTATGTGGCATCGAA	2312
Qy	741	ARGGLIMETGINIIEPRO 746	
Db	2313	AGAGAAATGAGAAATCCCT 2330	
RESULT 7			
LOCUS	BD186247		
DEFINITION	START6 activating gene.	2558 bp	DNA
ACCESSION	BD186247		linear
VERSION	BD186247.1		
KEYWORDS	WO 0206943-A/116.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
	Homidae; Homo.		

REFERENCE	1 (bases 1 to 2558)
AUTHORS	Honda, G., Matsuda, A., Muramatsu, S. and Ishizawa, K.
TITLE	State activating gene
JOURNAL	ASAI KASEI CORP, GOICHI HONDA, AKIO MATSUDA, SHUJI MURAMATSU, KENYA ISHIZAWA
COMMENT	OS Homo sapiens (human) PN WO 0206943-A/116 PD 05-DEC-2002 PF 22-MAY-2002 WO 2002JP004949 PR 25-MAY-2001 JP 01P 157043, 30-AUG-2001 JP 01P 260681 PR ° 10-OCT-2001 JP 01P 313175 PI GOICHI HONDA, AKIO MATSUDA, SHUJI MURAMATSU, KENYA ISHIZAWA PC C07K14/47, C07K17/00, C12N15/12, C12N5/10, C12P21/02, C12P21/08, PC C12Q1/02, PC A61P3/06, A61P3/10, A61P29/00, A61P31/00, A61P35/00, A61P37/00, PC G01N33/15, PC G01N33/50, G01N33/53, G01N33/566, G06F17/30, G06F17/60 CC STATE activating gene FH Key Location/Qualifiers FT CDS (58) . . (2295) . Location/Qualifiers 1. .2558 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
ORIGIN	
Alignment Scores:	
Pred. No.:	0
Score:	4038.00
Percent Similarity:	99.7%
Best Local Similarity:	99.7%
Query Match:	99.7%
DB:	2
US-10-773-302-2 (1-746) x BD186247 (1-2558)	
QY	1 MetGlyGlnThrGlyLysLeuSerGluValProValCysTrpAlaGlyValGValLys 20
DB	58 ATGGGCGAAGCTGGGAAGAAATCTGAAGAGGAGCAAGCTTGTGGCGGAAGCGGTAAAA 117
QY	21 SerGlnTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40
DB	118 TCGAGATGACATGGACATGACAGACCTCAAGAGTTCAACAGACGATGAGAAATGAAGT 177
QY	41 MetPheSerSerAsnArgGlnLysIleLeuGluArgTrpGlnIleLeuAsnGlnTyr 60
DB	178 ATGTTTAGTTCCAAATGTCAGAAATTTTGGAAAGAACGAAATCTTAACCAAGAAATGG 237
QY	61 LysGlnArgArgGlnLeuProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80
DB	238 AAACAGGAAGATATCAGCGCTGTGACATCTGACTTCTGAGAGCTCATTCGCGGAGCT 297
QY	81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100
DB	298 AGGAGAGTTCGGGTGACAGTACTGGAATTTTCCAAACAAGCATCCACTTAAGACT 357
QY	101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnAsnPhe 120
DB	358 CTGAATGCAAGTTCCTTCAATGACCACTAATGATTTCTTGATCTCCCTCAGCAGAAATTTT 417
QY	121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140
DB	418 ATGGTGAAGATGAACCTGTTTATCAATCAATCTTATATGGGAGATGAAGTTTATGAT 477
QY	141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyValValHisGlyAsp 160
DB	478 CAGAGATGATCTTTCATTGACAACTAATAAAAATTTGATGGGAAAGATACACGGGAGAT 537
QY	161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGln 180
DB	538 AGGAATATGGGTTTATATATGAAATTTTGTGGAGTTGGTGAATGCCCTTGGTCAA 597



Query Match: 99.7% Indels: 0  
DB: 2 Gaps: 0  
US-10-773-302-2 (1-746) x AX821950 (1-2576)

QY 1 MetGlyGlnThrGlyValLeuSerGlnValGlyProValCysTrpArgValLeuValLys 20  
DB ATGGGCCAGACTGGAGAGAAATCTGAGAGAGAACCAAGTTGTGGCGGAGACGGTGAATA 117  
QY 21 SerGlyTrpMetCysLeuArgGlnLeuLysArgPheArgAlaAspGlnValLysSer 40  
DB 118 TCAGAGTACATCCGACATGAGACAGCTCAAGAGTTTCAGACAGAGCTGAGAGTAAGAGT 177  
QY 41 MetPheSerSerAsnArgGlnLysLeuGlnValArgThrGlnLysLeuSerGlnGluTrp 60  
DB 178 ATGTTAGTTCCAAATCGTCAGAAAATTTGGAAAAGAACGAAATCTTAAACCAAGAAATGG 237  
QY 61 LysGlnTrpArgAlaGlnProValHisLeuThrSerValSerSerLeuArgGlyThr 80  
DB 238 AAACAGCGAAGATACAGCTGTGCAATCTGACTTGTGAGCTCATTCGCGGAGACT 297  
QY 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValLysProLeuLysThr 100  
DB 298 AGGAGTCTTCGGTGCACAGTCACTTGATTTCCACACAGTCAATCCATTAAAGACT 357  
QY 101 LeuAsnAlaValAlaSerValProLysMetTrpSerTrpSerProLeuGlnAsnPhe 120  
DB 358 CTGAATGCAATGGCTTCAGTACCCCAATATGTAATCTTGCTCCCTCAGACAGAAATTTT 417  
QY 121 MetValGlnAspGlnThrValLeuHisAsnLysProTrpMetGlnAspGlnValLeuAsp 140  
DB 418 ATGTGTGAAGATGAAATCTGTTTACATAACATCTCTTATATGAGAGATGAAGTTTAGAT 477  
QY 141 GlnAspGlyThrPheLeuGlnLysLeuLysAsnTrpAspGlyLysValHisGlyAsp 160  
DB 478 CAGATGTGACTTTTCATTGAAGAATATAAAAAATTATGATGGAAAGTACACGGGAGT 537  
QY 161 ArgGlnCysGlyPheLeuAsnAspGlnLysPheValGlnLeuValAsnAlaLeuGln 180  
DB 538 AGAAGATGTGGGTTTAAATGATGAATTTTGGAGTGTGGAATGGCCCTTGCTGACAA 597  
QY 181 TyrTrpAspAspAspAspAspAspGlyAspAspProGlnGlnValArgGlnLysGln 200  
DB 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 658 AAAGTCTGGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 717  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGlnLeu 240  
DB 718 GATAAATATTTTGGAGGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGACTA 777  
QY 241 LysGlnLysTrpLysGlnLeuThrGlnGlnGlnLeuProGlnValAlaLeuProGlnLys 260  
DB 778 AAGAAAAATATTAAGAACTACCGAAGACAGCTCCAGCGGCACTTCTCCGTAAGT 837  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer 280  
DB 838 ACCCCCAATGATGATGACCAATGCTAAATCTGTCAGAGAGCAAAAGCTTACACTCC 897  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisProPheHis 300  
DB 898 TTTCAATAGCTTTTCTGTAGCGATGTTTAAATGTGACTGCTTCTCAATCTTTTCAAT 957  
QY 301 AlaThrProAsnThrTrpLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB 958 GCAACACCAACCACTTATAGCGAGAAACACAGAAACAGCTCTAGACAAACCAACCTTGT 1017  
QY 321 GlnProGlnCysTrpGlnHisLeuGlnGlnValLysGlnPheAlaAlaLeuThrAla 340  
DB 1018 GGACCAAGTGTACCAAGCAATTTGGAGGGAGCAAGAGTTTGTCTGCTCTCACCGCT 1077

QY 341 GlnArgIleLysThrProProLysArgProGlnGlyValArgArgGlyArgLeuProAsn 360  
DB 1078 GAGCGATTAAGACCCCAACAAACCTCCAGAGGCGCGCAAGAGGAGCGCTTCCAAAT 1137  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1138 AACGATGACAGGCCACAGCACCCCACTTAATGTCTGAAATCAAAAGATACAGACAGT 1197  
QY 381 AspArgGlnAlaGlyThrGlnThrGlyGlnLeuAsnAspLysGlnGlnGlnLys 400  
DB 1198 GATAGGAGAACAGAGGATGAAACGGGGGAGAGAACATATTAAGAAAGAAAGAAAG 1257  
QY 401 LysAspGlnThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProLysMet 420  
DB 1258 AAAGATAAATCTGACCTCTCTGAAGCAAAATCTCGGTGTCAACACCAATTAAGATG 1317  
QY 421 LysProAsnIleGlnProProGlnAsnValGlnTrpSerGlnAlaGlnLysMetPhe 440  
DB 1318 AAGCCAAATTTGAACCTCTGAGAAATGTGAGAGTGTGTGTGTAAGCTCAATGTTT 1377  
QY 441 ArgValLeuIleGlyThrTrpTrpAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1378 AGAGTCTCAATGGCACTTACTATGACAATTTCTGTCTCATTGTAGTTAAATGGGAGC 1437  
QY 461 LysThrCysArgGlnValTrpGlnPheArgValLysGlnSerIleIleAlaProAla 480  
DB 1438 AAAACATGTAGACAGGTATGAGTTTAAAGTCAAAACCTGATCATACATAGCTCCAGCT 1497  
QY 481 ProAlaGlnAspValAspThrProProArgLysLeuLysArgLysHisArgLeuTrpAla 500  
DB 1498 CCCGCTAGATGTGATCTCTCTCCAGAGAAAGAAAGAGAAACCGGTTGTGGGCT 1557  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTrpAsnTrp 520  
DB 1558 GCACACTGCAAGAAAGATACAGCTGAAAAAGACGGCTCTCTAAACATGTTTCAACATAT 1617  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1618 CAACTGTATATCATCCACAGCGACCTTGACAGTGTGACCTTGTGTATGACAA 1677  
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlnCys 560  
DB 1678 AATTTTGTGAAAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
QY 561 ArgCysAlaIleGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValArgGlnCys 580  
DB 1738 CGCTGCAAGACAGTGCACACCAACAGAGCCCGGTCTACCTGCTGCTCGGAGATGT 1797  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisThrPheSerLysAsnValSer 600  
DB 1798 GACCTTACCTGTGTTACTTGTGAGCGGTGACCAATGGGACAGTAAATAATGTGTCC 1857  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAsp 620  
DB 1858 TGCAGAACTGCAATATTCAGCGGGGCTCCAAAAGATCTATTTGCTGGCAGCATCGAC 1917  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGln 640  
DB 1918 GTGGCAGGCTGGGGGATTTTATATCAAGATCTGTGCAAGAAATGAATCATCTCAGAA 1977  
QY 641 TyrCysGlyGlnIleIleSerGlnAspGlnAlaAspArgArgGlyLysValTrpAspLys 660  
DB 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGAATAA 2037  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
DB 2038 TACTGTGCACTTCTGTTCACTTGAACATATATTTGTGTGTGATGACACCGGAG 2097  
QY 681 GlnLysLysIleArgPheAlaAsnHisSerValAsnProAsnCysTrpAlaLysValMet 700  
DB 2098 GGTAAACAAATTTGTTTGAAGATCATTCGGTAATCAAACTGCTATGCAAAAGTTATG 2157  
QY 701 MetValAsnGlnAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGln 720

Db 2158 ATGGTTACGGTGCATCAGCATAGTATTTTGGCAAGAGCCATCCAGCTGGCGAA 2217  
Qy 721 Gtuleuphepneapstytatgttyserginalaapalaleuylsttvaly11eglu 740  
Db 2218 GAGCTGTTGTTGATTACAGATACAGCCAGCTGATGCCCTGAAGATATGCGCATCGAA 2277  
Qy 741 ArgglumetgluilePro 746  
Db 2278 AGAGAAATGAAATCCCT 2295

RESULT 9  
LOCUS HSZESTENH 2576 bp mRNA linear PRI 18-APR-2005  
DEFINITION H.sapiens mRNA for enhancer of zeste.  
ACCESSION X95653  
VERSION X95653.1 GI:1438063  
KEYWORDS zeste enhancer gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 2576)  
Chen H., Rosier C. and Antonarakis S. E.  
Cloning of a human homolog of the Drosophila enhancer of zeste gene  
(EZH2) that maps to chromosome 21q22.2  
Genomics 38 (1), 30-37 (1996)  
8954776  
2 (bases 1 to 2576)  
Antonarakis S. E.  
Direct Submission  
Submitted (13-FEB-1996) S. E. Antonarakis, Division of Medical  
Genetics, University of Geneva School of Medicine, Rue  
Michel-Servet 1, CH-1211 Geneva, SWITZERLAND  
Related protein P42124.  
Location/Qualifiers  
1..2576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="q22.2"  
/tissue\_type="Brain"  
/dev\_stage="Fetal"  
1..2576  
/gene="EZH2"  
58..2298  
/gene="EZH2"  
/codon\_start=1  
/product="enhancer of zeste"  
/protein\_id="CAA64955.1"  
/db\_xref="GI:1438064"  
/db\_xref="GOA:O15910"  
/db\_xref="UniProtKB/Swiss-Prot:Q15910"  
/translation="MGQTGKSEKPVCMWRKVSRYRLNQLKFRPADYKMFSS  
NRKILRTIELINQKQRIQPVHILTVSLKTRCSTVSDLPFOYIPLKTLN  
AVASVPIWMSPLQONFVDETVLANIPIVGDVLDQDFIIEELIKNDGKYGHD  
RECGFINDEIPIVELNALGOYVNDDDDDGDPERBERKQLEBHDRODKSPRPF  
PSDKLEIASSWFPDKGTABELKEKYKELTEQQLPGALPPECTPIIDGPNKSVYRQ  
SLHSPTLIPCRCPYDCFLPHPHATPTTYRKATETALDNKPCPGCYOYLKAKER  
AALTAERIKTPKPGKRRKRLFNNSRSTPTIVLESKDTDSDEAGETGGEN  
NDKEBERKDETSSEANSRCQPIIKKPILEPENVWSGASMPRLVIGTYDN  
FCALAIRLIGTKICQVYFRVSESIIAPAPADVDTPRKRRGRLMAHCRKIQL  
KKDSNVYVYQPCDHROPDCSSPCVIAONFCERKQSSSCONFPKGRCAQC  
NTKQPCYLAABCDPDLCTGCAADHDSKNSCKNGSIORGSKHLLAPSDVAG  
GIPTDPOKNERISEYGEIISQDEADBRGVYKWSCELFNLPNDPVUDAPKKN  
KIRFANHSVNPNCIAKMMVNGDHRIGIPAKAIDTGSELFDVIRYSQADMLKTVGIS  
REMERIP"  
2558  
/gene="EZH2"  
polYA\_site  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 2  
Query Match: 99.7% Indels: 0  
Gaps: 0  
US-10-773-302-2 (1-746) x HSZESTENH (1-2576)  
Qy 1 Metgljgnrnrglylyeysersglulysgliprovalcyetparylysarvallys 20  
Db 58 ATGGCGCAGACTGGGAAGAAATCTGAGAAGGACAGAGTTGTTGGCGGAAGGTGTAATA 117  
Qy 21 Sergluylmetargleuargluleuylsarpleaarglalaargluvalylsser 40  
Db 118 TCAGAGTACATCGCATGACAGACGTCAAGAGTTCACACGAGCTGAAGATGAAGT 177  
Qy 41 MetPheSerSerAspargluleylleuugluargthrngluileuansglnlurtP 60  
Db 178 ATGTTAGTTCATGTCGAGAAATTTGGAAAGACGAAATCTTAACCAAGATGG 237  
Qy 61 LysglinararglileglnprovalhisileuThrSerValSerSerLeuargllythr 80  
Db 238 AAACGCGAAGGATACAGCTGTGACATCTTGAGCTCATTTGGCGGAGCT 297  
Qy 81 ArgglucyservalthrsarpleuapPheprothrglnvalilProleuylthr 100  
Db 298 AGGAGTGTTCGGTACCAAGTGAATTTCCAAACAAGTATCCCATTAAGACT 357  
Qy 101 Leuana1av1alaserValProleuylSerThrSerProleuuglnanpPe 120  
Db 358 CTGAATGACAGTGTCTTCACTGACACCAATAGATTTGTGCTCCCTTACAGCAATTTT 417  
Qy 121 MetvalgluaargluthrValleuHisanileprolyrmetgluYaspgluValleuap 140  
Db 418 ATGGTGAAGATGAAGACGTTTACATTAACATCTCTTAATGCGAGATGAAGTTTAAAT 477  
Qy 141 GlnaarglythrPheileglugluuileuylsabyrYrarglylyvalHisgluYasp 160  
Db 478 CAGATGTGACTTTCATTAAGAACTAATAAAATTTATGATGGGAAAGTACACCGCGAT 537  
Qy 161 Argglucysgllypheleleuanspgluilephevaligluuvalanala1euglygln 180  
Db 538 AGAGATGTGGGTTTAATAATGATGAATTTTGTGGAGTGGGTAATGCCCTTGCTATA 597  
Qy 181 TyranaBPAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
Db 598 TTTAATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
Qy 201 LysaBPleuugluasphilaYrAspAspAspAspAspAspAspAspAspAspAspAsp 220  
Db 658 AAAAGTCTGAGAGATCACCGAGATGATAAGAAAGCCGCCACCTCGGAAATTTCTTCT 717  
Qy 221 AspylllePhegluala11eserSerMetPheProaBPlysglythrAla1glnluu 240  
Db 718 GATAAATTTTGGAGGCCATTTCTCAATGTTTCCAGATTAAGGGACAGCAAGAAACTTA 777  
Qy 241 LysgluylsYrlysgluuethrGlnuglnleuProglYalaleuProProgluYs 260  
Db 778 AAGGAAAAATTAAGAACTACCGAAGACGCTCCAGGCGCATCTCTCTGTAATGT 837  
Qy 261 ThrProaenlleaarglyProaana1alyservalGlnarggluGlnserleuHisser 280  
Db 838 ACCCCAAACATAGATGACAAATGCTAAATCTGTTGAGAGAGCAAGCTTACATCC 897  
Qy 281 PheHisThrleuPheCysArxArgCysPheleYsYrArpCysPheleuHisProPheHis 300  
Db 898 TTTCAATAGCTTTTCTGTAGCGAGATGTTTAATATGATGCTCTTCTTCAATCTTTTAT 957  
Qy 301 AlaThrProaenThrYrlysaYrlysaYrlysaYrlysaYrlysaYrlysaYrlysaYr 320  
Db 958 GCAACACCAACATCTTAATACCGGAGAGACAGAAACAGCTTATGACAAACCTTGT 1017

Qy 321 G1yProGlnCybYrGlnHleuGlnG1yAlayGlnPhealaa1aleuThrAla 340  
 Db 1018 GGACCAAGGTGTTCCAGCATTTGGAGGGAGCAAGAGGTTGGCTGCTCACCCT 1077  
 Qy 341 G1uArg11elyThrProProlybArgProG1yG1yArgArgG1yArgLeuProasn 360  
 Db 1078 GAGGGAGTAAAGACCCCAACAAAGTCACAGAGCCGACAGAGAGAGCGCTCCCAT 1137  
 Qy 361 AasnSerArgProSerThrProThr11easnVal1leuGlnSerLybAspThrAspSer 380  
 Db 1138 AACGATGACAGCCCAAGACCCCAACATTAATGCTGCGAATCAAGAGATCAAGACGT 1197  
 Qy 381 AspaArg1ualag1yThrG1yG1yG1yG1yG1yG1yG1yG1yG1yG1yG1yG1y 400  
 Db 1198 GATGGAGAGCAGGGAGCTGAACCGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1257  
 Qy 401 LybAspG1uThrSerSerSerSerSerG1uAlaAsnSerArgG1yG1yG1yG1yG1y 420  
 Db 1258 AAAGATGAACTTGAGCTCTCTGAGCAAAATTCGCTGTCAACACCAATTAAGATG 1317  
 Qy 421 LybProasn11eg1uProProG1yAsnVal1G1uTPSerG1yAlaG1uAlaSerMetPhe 440  
 Db 1318 AACCAATATTTGAACCTCTCGAATATGAGATGAGATGCTGCTGAAGCTCAATGTT 1377  
 Qy 441 ArgVal1leu11eg1yThr1yTrAspaAspPheCybAla11eal1aArgLeu11eg1yThr 460  
 Db 1378 AGAGCTCTCATTTGAGCATTAATGACATTTCTGTGCTAGCTAGGTAAATGGAGCC 1437  
 Qy 461 LybThrCybArgGlnVal1yTrG1uPheArgVal1yG1yG1yG1yG1yG1yG1yG1y 480  
 Db 1438 AAAACATGTACACAGGTATGATGATTAAGCTCAAAAGATTAACATCATGTGCTCAGCT 1497  
 Qy 481 Pro1aG1uAspVal1AspThrProProArg1yLybLybLybLybLybLybLybLybLyb 500  
 Db 1498 CCCCTGAGATGATGATGATCTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557  
 Qy 501 Alah1bCybArg1y11eg1uLeu1yLybAspG1ySerSerSerAsnH11eal1yAsn1y 520  
 Db 1558 GCACACATGCAAAAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 Qy 521 G1nProCybAspH11AspArgG1nProCybAspSerSerCybProCybVal11eal1Gln 540  
 Db 1618 CAACCTGTGATCACTCAACGAGAGCTGTGACAGTTGCTGCTGTGATAGACAA 1677  
 Qy 541 AasnPheCybG1uLybPheCybG1nCybSerSerSerG1uG1nAsnArgPheProG1yCyb 560  
 Db 1678 AATTTTGTGAAAAGTTTGTCAATGTATGTTCAAGATCAAAACCGCTTCCGGAGTC 1737  
 Qy 561 ArgCybLybAlaGlnCybAsnThrLybGlnCybProCybTyTrLeuAlaVal1ArgGluCyb 580  
 Db 1738 CGTGCAGAAAGACAGTGCAGACCAAGAGAGTGCCTGCTGCTGCTGCTGCTGCTGCT 1797  
 Qy 581 AspProAspLeuCybLeuThrCybG1yAla11aAspH11eal1aAspH11eal1aSer 600  
 Db 1798 GACCTGACCTCTGCTTACTTGTGAGACCCCTGACCACTTGAGAGAGAGAGAGAGAG 1857  
 Qy 601 CysLybAsnCybSer11eg1nArgG1ySerLybLybLybLybLybLybLybLybLyb 620  
 Db 1858 TGCAGAGATCTCAATATTCAGCGGGGCTCCAAAAGCATTAATGCTGCGACCATCTGAC 1917  
 Qy 621 Val1aG1yTrpG1y11ePhe11elybAspProVal1G1uLybAsnG1uPhe11eSerG1u 640  
 Db 1918 GTGGAGAGCTGGGGAGTTTATTAACAATCTGTCGAGAAATGAATTCATCTCAGAA 1977  
 Qy 641 TyTrcG1yG1u11e11eSerG1nAspG1uAlaAspaG1yG1yLybVal1yTrAspLyb 660  
 Db 1978 TACTGTGAGAGATTAATTTCTCAAGATAGACCTGCAGAGAGAGAGAGAGAGATGAATA 2037  
 Qy 661 TyTrcSerSerPheLeuPheAsnLeuAsnAsnAspPheVal1Val1AspAla1ThrArgLyb 680  
 Db 2038 TACATGTGAGCTTCTGTCACTTGAACATGATTTGTGTGAGATGCAACCGCAGAG 2097

Qy 681 G1yAsnLyb11eArgPheAlaAsnH11eSerVal1AsnProasnCybTyTrAlaLybValMet 700  
 Db 2098 GGTAAACAAATTCCTTTGCAAAATCATTCGTAATCAAACTGCTATGCAAAAGTTATG 2157  
 Qy 701 MetVal1asnG1yAspH11eArg11eg1y11ePheAlaLybArgAla11eg1nThrG1yG1u 720  
 Db 2158 ATGTTTAACTGATATCAAGATATGATATTTTTCACAGAGAGAGAGAGAGAGAGAGAG 2217  
 Qy 721 G1uLeuPhePheAspTyTrArgTyTrSerG1nAlaAspAla1leuLybTyTrVal1G1yG1u 740  
 Db 2218 GAGCTGTTGTGTATTAACATATACAGCAGCAGATGCTGCTGGAATATGTGCGATCGAA 2277  
 Qy 741 ArgG1uMetG1u11ePro 746  
 Db 2278 AGAGAAATGAGAAATCCCT 2295

RESULT 10  
 LOCUS BC010858  
 DEFINITION Homo sapiens enhancer of zeste homolog 2 (Drosophila), transcript variant 1, mRNA (cDNA clone MGC:9169 IMAGE:3901250), complete cds.  
 ACCESSION BC010858  
 VERSION BC010858.2 GI:34194096  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
 AUTHORS 1 (bases 1 to 2655)  
 Strausberg, R.L., Peingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.U., Uedini, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.U., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.D., Malek, D.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hu, Y.K., S.M., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schnerch, A., Schein, J.B., Jones, S.J. and Marra, M.A.  
 Mammalian Gene Collection Program Team  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 2655)

CONSRM  
 TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 CONSRM  
 JOURNAL  
 REMARK  
 COMMENT

NIH MGC Project  
 Direct Submission  
 Submitted (12-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 On Aug 25, 2003 this sequence version replaced gi:14790028.  
 Contact: MGC help desk  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUMI)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-sbnc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.



QY 456 ArgLeuIleGlyThrIleThrCysArgGlnValIleThrGluPheArgValIleGluSerSer 475  
DB 1503 AGTTAAATGGGACCAAAACATGATGACAGGTGATGAGTTTAAAGTCAAAAGATCTAGC 1562  
QY 476 IleIleLeuAlaProAlaProAlaGluAspValAspThrProAlaGlyIleValAspArg 495  
DB 1563 ATCTAGTCTCAGCTCCCGTGAAGAGTGTGATCTCTCCAAAGAAAAGAGAGAGAA 1622  
QY 496 HisArgLeuTrpAlaAlaHisCysArgGlyIleGlnLeuIleValAspGlySerSerAsn 515  
DB 1623 CACCGGTTGGGCTGACACCTGCAGAAAGATACAGCTGAAAAGAGAGCGCTCTCTAAC 1682  
QY 516 HisValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 535  
DB 1683 CATGTTTACACACTATCAACCTGTGATCATCCAGCGAGCTTGTGACAGTTCGTGCCCT 1742  
QY 536 CysValIleIleAlaGlnAsnPheCysGlnIleValPheCysGlnIleValIleValIleVal 555  
DB 1743 TGTGTGATGACCAAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAAC 1802  
QY 556 ArgPheProGlyCysArgCysValAlaGlnCysAsnThrIleValIleValIleValIleVal 575  
DB 1803 CGCTTCCGGATGCGCTGCAAGACAGTGCACACCAAGCAGTGCCTGCTACTG 1862  
QY 576 AlaValArgGluCysAspProAspIleuValIleValIleValIleValIleValIleVal 595  
DB 1863 GCTGTCCGAGAGTGAACCTGACCTGTCTTACTTGTGAGCGCTGACCAATGGGAC 1922  
QY 596 SerIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 615  
DB 1923 AGTAAATATGTCTCTGCAAGAACTGCAAGTATTCAGCGGCTCCAAAGAGATCTATTG 1982  
QY 616 LeuAlaProSerAspValAlaGlyTrpGlyIlePheIleValAspProValGlnIleValAsn 635  
DB 1983 CTGCAACCATCTGACGTGAGGCTGGGGATTTTATCAAAGATCCGTGCAAGAAAT 2042  
QY 636 GluPheIleSerGluTrpCysGlyGlnIleIleSerGlnAspGlnIleAspArgArgGly 655  
DB 2043 GAATTCATCTCAGATATCTGTGAGAGATTTTCTCAAGATGAAGCTGACAGAAAGGG 2102  
QY 656 LysValIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 675  
DB 2103 AAAGGTATGATTAATATACATGTCAGCTTCTGTTCATCTGAACAAAGATTTTGTGTG 2162  
QY 676 AspAlaThrArgGlyValAsnIleValIleValIleValIleValIleValIleValIleVal 695  
DB 2163 GATGCAACCCCGAAGGATACAAATTCGTTTGCATTCATTCGTTAAATCCAAACCTGC 2222  
QY 696 TyrAlaIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 715  
DB 2223 TATCAAAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2282  
QY 716 IleGlnThrGlyGluGlnLeuPhePheAspIleValIleValIleValIleValIleValIleVal 735  
DB 2283 ATCCAGACTGCGGAGAGAGCTGTTTGTGATTAACAGATACAGCAGGCTGATGCCCTGAG 2342  
QY 736 TyrValGlyIleGluArgGluMetGluIlePro 746  
DB 2343 TATGTCCGATCGAAAGAGAAATGAATTCCT 2375

RESULT 11  
DD208658 2512 bp DNA linear PAT 19-JAN-2006  
LOCUS Expression Profile of Prostate Cancer.  
DEFINITION DD208658  
ACCESSION DD208658.1 GI:85649860  
VERSION JP 200518522-A/95.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2512)

AUTHORS Rabin, M.A., Chinaiyan, A.M. and Surikumar, A.  
TITLE Expression Profile of Prostate Cancer  
JOURNAL Patent: JP 200518522-A 95 23-JUN-2005;  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN  
COMMENT OS Homo sapiens  
PN JP 200518522-A/95  
PD 23-JUN-2005  
PF 02-AUG-2002 JP 2003517245  
PR 02-AUG-2001 US 60/309581, 15-NOV-2001 US 60/334468, PR  
01-AUG-2002 US 10/210120  
PI mark a rabin, arun m chinaiyan, arun surikumar CC  
FH key Location/Qualifiers  
FEATURES  
source 1..2512  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2512  
Score: 4024.00 Matches: 745  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.3% Indels: 1  
DB: 2 Gaps: 0  
US-10-773-302-2 (1-746) x DD208658 (1-2512)  
QY 1 MetGluValIleThrGlyIleValIleValIleValIleValIleValIleValIleValIleVal 20  
DB 10 ATGGGCGGAGCTGGAGAAATCTGAGAGAGGACCAAGTTGTGGCGGAGCGTGTAAA 69  
QY 21 SerGluIleThrArgLeuArgGlnIleValIleValIleValIleValIleValIleValIleVal 40  
DB 70 TCAGAGTACATCGACATGAGCAGCTCAGAGGTTCAAGAGCTGATGAAAGTAAAGT- 128  
QY 41 MetPheSerSerAsnArgGlnIleValIleValIleValIleValIleValIleValIleValIleVal 60  
DB 129 ATGTTTGTTCATTCCTCAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 188  
QY 61 LysGlnIleThrArgGlnIleValIleValIleValIleValIleValIleValIleValIleValIleVal 80  
DB 189 AAACAGGAGGAGATACAGCTGTGACATCTCTGTGACCTCATTTGCGGGAGCT 248  
QY 81 ArgGluCysSerValIleThrSerAspLeuAspPheProThrGlnIleValIleValIleValIleVal 100  
DB 249 AGGAGTGTTCGGTGAACAGTCACTGATTTTCCAAACAGATCCCATTAAGACT 308  
QY 101 LeuAsnAlaValAlaSerValProIleMetIleSerIleTrpSerProLeuGlnIleAsnPhe 120  
DB 309 CTGAATCAGTGTCTCAGTACATCATATGATTTCTGTGCTCCCTTACAGCAGAAATTTT 368  
QY 121 MetValGluIleAspGluIleValIleValIleValIleValIleValIleValIleValIleValIleVal 140  
DB 369 ATGTGTGAAGATGAACGTTTATACATTAACATTTCTTATATGGAGATGAAGTTTATGAT 428  
QY 141 GlnAspGlyThrPheIleGluGlnIleValIleValIleValIleValIleValIleValIleValIleVal 160  
DB 429 CAGATGTGACTTTCATTTGAAGAACTATTAATAAAATTTATGATGGAGAAAGTACAGGGGAT 488  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaIleGlyGln 180  
DB 489 AGAAGATGTGGGTTATTAATGAGAAATTTTGTGAGTGTGAGAAATGCTTGTGTGCA 548  
QY 181 TyrAsnAspAspAspAspAspAspGlyIleAspAspProGluGluIleValIleValIleValIleVal 200  
DB 549 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608  
QY 201 LysAspLeuGluAspIleAspAspAspIleValIleValIleValIleValIleValIleValIleVal 220  
DB 609 AAAGATCTGAGAGATCAACGAGATGATTAAGAAAGCCGCGCACTCGGAAATTTCTTCT 668



221 AaplylePheGlu1a1a1eSerSerMetPheProAaplyeGlyThra1a1a1eGlu1e 240  
 669 GATAAATTTTGAAGCCATTTCTCATGTTTCCAGATAGGCGACACAGAAAGACTA 728  
 241 LygGlyuYrYrGlu1eThrgu1eThrgu1eThrgu1eThrgu1eThrgu1eThrgu1e 260  
 729 AAGGAAAAATATTAAGAACTACCGAAACAGAGCTCCAGCGCACTTCTCTGATGT 788  
 261 ThrProen1leapGlyProaen1a1ySerValGlna1a1eGlu1eThra1a1e 280  
 789 ACCCCCACTATGATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 848  
 281 PheH1eThra1ePheCysa1a1eThra1eThra1eThra1eThra1eThra1eThra1e 300  
 849 TTTCTAGGCTTTCTGTAAGCGATGTTTAAATGAGCTGCTCTTACATCTTTTCT 908  
 301 AlaThra1eProaen1Thra1eThra1eThra1eThra1eThra1eThra1eThra1e 320  
 909 GCAACCCCACTATGATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 968  
 321 GlyProGlnYrYrGlu1eThrgu1eThrgu1eThrgu1eThrgu1eThrgu1eThrgu1e 340  
 969 GGACCAAGTGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1028  
 341 GluA1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 360  
 1029 GAGCGATTAAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1088  
 361 AenSerSerArgProSerThra1eThra1eThra1eThra1eThra1eThra1eThra1e 380  
 1089 AACAGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1148  
 381 AapA1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 400  
 1149 GATAGGAGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1208  
 401 LysAapGluThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 420  
 1209 AAGAGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1268  
 421 LysProaen1eGluProGlu1eThra1eThra1eThra1eThra1eThra1eThra1e 440  
 1269 AAGCAATATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1328  
 441 ArgVal1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 460  
 1329 AAGAGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1388  
 461 LysThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 480  
 1389 AAGCAATATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1448  
 481 Proa1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 500  
 1449 CCGCGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1508  
 501 AlaH1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 520  
 1509 GACAGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1568  
 521 GlnProCysAapH1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 540  
 1569 CACCGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1628  
 541 AapH1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 560  
 1629 AATTTTGAAGAAATTTTGAAGAAATTTTGAAGAAATTTTGAAGAAATTTTGAAGAA 1688  
 561 ArgCys1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 580  
 1689 CGGTGCAAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAG 1748  
 581 AapProa1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 600

1749 GACCTGACCTCTGCTACTTGTGAGACCCGTGACCATTTGGAGACATGAAATGTC 1808  
 601 Cys1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 620  
 1809 TGCAAGAACTGAGTATGAGCAATGCTATGCTGAGAGAGCAAGGCTTACACTCC 1868  
 621 Val1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 640  
 1869 GTGGAGGCTGAGGATTTTATGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 1928  
 641 TyrCys1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 660  
 1929 TACTGTGAGAGATTTTATGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 1988  
 661 TyrMetCysSerPhe1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 680  
 1989 TACATGTGACCTTTCTGCTCAATGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 2048  
 681 GlyAen1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 700  
 2049 GGTAACTAAATGCTTGTGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 2108  
 701 MetVal1a1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 720  
 2109 ATGCTTACGAGTATGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 2168  
 721 Glu1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 740  
 2169 GAGCTGTTTGTGATTAAGTATGAGCAATGCTATGAGAGAGCAAGGCTTACACTCC 2228  
 741 ArgGlu1eThra1eThra1eThra1eThra1eThra1eThra1eThra1eThra1e 760  
 2229 AAGAAATGAGAAATCCCT 2246

RESULT 12  
 BC003772 2574 bp mRNA linear ROD 29-JUN-2004  
 LOCUS Mus musculus enhancer of zeste homolog 2 (Drosophila), mRNA (cDNA  
 DEFINITION  
 accession BC003772  
 version BC003772.1 GI:13277755  
 keywords MGC.  
 source Mus musculus (house mouse)  
 organism Mus musculus  
 reference Strausberg, R., L., Feng, G., D., Grouse, L. H., Derge, J. G.,  
 1 (bases 1 to 2574)  
 1 (bases 1 to 2574)  
 Strausberg, R., L., Feng, G., D., Grouse, L. H., Derge, J. G.,  
 Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D.,  
 Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bat, N. K.,  
 Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L.,  
 Stepieton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L.,  
 Schaefer, T. B., Brownstein, M. J., Ueda, T. B., Tohbi, Y. K., S.,  
 Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J.,  
 Abramson, R. D., Mullaly, S. J., Bosak, S. A., McEwan, P. J.,  
 McEwan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S.,  
 Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hui, Y. S. W.,  
 Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbe, R. A.,  
 Raney, J., Heiton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Shenchenko, Y.,  
 Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M.,  
 Butlerfield, Y. S., Krzywinski, M. J., Skalski, U., Smalls, D. B.,  
 Schnerch, A., Schein, J. B., Jones, S. J., and Marra, M. A.,  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 JOURNAL  
 PROC. NATL. ACAD. SCI. U.S.A. 99 (26), 16899-16903 (2002)  
 PUBMED  
 12477932  
 REFERENCE  
 2 (bases 1 to 2574)  
 Strausberg, R.



**TITLE**  
**JOURNAL**

**Direct Submission**  
Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shmc.stanford.edu>  
 Contact: (Dickson, Mark) [mcdeparill@stanford.edu](mailto:mcdeparill@stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 8 Row: p Column: 1  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6679720.

## FEATURES

```

source
1.2574
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="MGC:5969 IMAGE:3492110"
/tissue_type="mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clone1) outgrowth infected with the virus MMTV."
/clone_1ib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1.2574
gene

```

**CDS**

/gene= b2a1  
 /codon\_start=1  
 /product="enhancer of zeste homolog 2"  
 /protein\_id="AAH03772.1"  
 /db\_xref="gi:13277756"  
 /db\_xref="GeneID:14056"  
 /db\_xref="WGI:107940"  
 /translation="WDGQTKGSKSEKPCVWRKRVKSYEMLRQKFKPRADVEKTMFSS  
 LKINSLATLQNMWQKRIQPHMTVSILGRBQSVSDLPFAQVPLTTL  
 AVASPIVMSVSPKQPMVEDEVDLITPYMDVEDDGFIEELINYGKPKH  
 RECGINDEIVELVNLNGVDNDDDDDDDDEBEEKQDLNDRDEKQ  
 PADTIKPRASMPDQKQTLAKPRKATLTLEQLPALPCCPIINDGNMAVS  
 SLSEHTFICSRCKTQCFIAKPAITNTYKRNITLADNKCPCQCYOCTHGAKEK  
 AALTALEIKTPPRPGGRGRGLPNNSVSPPTLVLESKDQDSDEAGETGESE  
 NDKEBKEKDESSSSSSSSSSQPIIMKRNIDVPBPNVMSGAEMFVILGYIDD  
 NCALAILGTCTGCVQVEFRKSESLIAPVTDIDVPKTKKRIKRLIAACRIQ  
 KDDSSQCLVARECDHPDRCSSGCVIAONKPCFOGSGECCNRRPGCAQNG  
 NTKQPCVLAARECDPDLCTGCAADHMSKVSCKNCSIRGSGKHLIAPSDVAG  
 GIFTDVPQKNEPISYCGEIIISDDEADRGKTYDYKMSGPIPLNANDVPDATKGG  
 IIRFANHSVNPCCAKVMWNGDRIIGIEMKRALIQGEELPDYRYSQNDATKYGII  
 IEMDELP"

**ORIGIN**

**Alignment Scores:**

Free. No.:	0	Matches:	734
Score:	399.00	Conservative:	7
Percent Similarity:	99.3%	Mismatches:	5
Best Local Similarity:	98.4%	Indels:	0
Query Match:	98.7%	Gaps:	0
DB:	6		

US-10-773-302-2 (1-746) x BC003772 (1-2574)

OY	1	MetGlyInThrgIyLylylSerGlnlyGlyProValCySerThrpaGlySerValyls	20
Db	54	ATGGGCCAGACTGGGAAAGAAATCGAAGAGGACCGGGTGGTGGCGGAAGCGTAA	113
OY	21	SerGlnTyMetCAsGleuArgGlnLeuIlyArgPheArgAlaSerGlnValylsSer	40
Db	114	TCAGAGTACATGAGACTGAGCAGCTCAAGAGGTTCCAGAGAGCTGATGAAGTAA	173
OY	41	MetPheSerSerAanaArgGlnlylLeuGlnArgThrgIulEleuamGlnlUTP	60
Db	174	ATGTTTGGTTCACATCGTCAGAAAATTTGGAAAGAACTGAACCTTAAACCA	233
OY	61	LyseGlnArgAglIegInProValHsIleleuthSerValSerSerLeuArgGlyThr	80
Db	234	AAGCAGCGGAGAAATACAGCTGTGCAATCATGACTCTGTGAGTCAATTGGCGG	293
OY	81	ArgGlnCySerValIThrSerAsPleuAsPheProThrgInValIleProleuylThr	100
Db	294	AGGAGGTGTTCACTCACACAGTGACTTGGAATTTTCCACACAAGTCAATCCG	353
OY	101	LeuAsnAlaValAlaSerValProIleMetCysSerThrSerProleuGlnIAsn	120
Db	354	CTGAATGAGTGGCTCGGTGCTTAATATGACTCTTGTCGCCCTTACACAGAA	413
OY	121	MetValGlnAspGlnThrValLeuHsAsnIleProTyMetGlyAspGlnValLeuAsp	140
Db	414	ATGGTGAAAGACGAAACGTCTTACATTAACATCTCTTACATGGGGATGAAGTT	473
OY	141	GlnAspGlyThrPheIleGlnGlnLeuIlelyAsnTyTrAspGlylyValHsGlyAsp	160
Db	474	CAGGACGGCACTTTCATTGAAAGAACTAATATAATATGATGGAAAGGCAATGG	533
OY	161	ArgGlnCySerGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlyGln	180
Db	534	AGAGATGTGATTTATTAATGATGAATTTTGTGAGAGTTGGTAAATGCTCTTG	593
OY	181	TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnIuaArgGlnlyGln	200
Db	594	TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	653
OY	201	LyAspArgLeuGlnAspPheIleArgAspAspLyseIuSerTrpProAlaGlyPheProSer	220
Db	654	AAAGATCTAGAGAAATCGAGATGAATAAAGAACTTGGCCACTCGGAATTTCTG	713
OY	221	AspLyseIlePheGlnAlaIleSerSerMetPheProAspLyseGlyThrAlaGlnIleu	240
Db	714	GATAAATATTTGGAAGCACTTTCCTCAATGTTTCCAGATAGGGCACCGCAGA	773
OY	241	LyseGlnlyGlyLylylLeuThrgIulGlnIleuProGlyAlaIleuProProGlnCys	260
Db	774	AAGGAAAAATTAAGAACTACAGAGAGCAGAGCTCCAGAGTGTCTGCTCCGAA	833
OY	261	ThrProAsnIleAspGlyProAsnAlaIlySerValGlnArgGlnIuSerLeuHsSer	280
Db	834	ACTCCAAACATCGATGAGCAAAATGCCAAATCTGTTCAGAGGAGCAAAAGCTT	893
OY	281	PheHsIThrLeuPheCysArgArgCysPheLyseTyTrAspCysPheLeuHsProPheHs	300
Db	894	TTTCAATAGCTCTGTGTCAGAGATGTTTAAGATGATGCTGCTCTTACATCC	953
OY	301	AlaThrProAsnThrTyLylylArgIlyAsnThrGlnThrAlaIleuAspAsnlyProCys	320
Db	954	GCAACACCCACACATATTAAGAGGAAGAACACAGAACAGCTTTGACACAAAG	1013
OY	321	GlyProGlnIuTyTrGlnHsIleuGlnIuAlaIlyGlnIuPheAlaAlaIleuThrAla	340
Db	1014	GGAACACAGTGTATACAGACTCTGAGAGGAGCTAAAGAGTTTGTGCTGCTTAC	1073
OY	341	GluArgIlyIlelylThrProProIuLyAspProGlyIyAlaArgAlaArgGlyIyArgLeuProAsn	360
Db	1074	GAGCGGATTAAGACACCACTTAACCGCCAGAGGGGCGCAGAAAGGAAGCTTCC	1133

```

QY 361 AanserSerArgProSerThrProThrIleAenValIleuGluSerIleAspThrAspSer 380
DB 1134 AACAGTAGAGACCCAGACCCACCAATCAGTGTGAGAGCAAGATTCAGACACT 1193
QY 381 AaPArgGluAlaGlyThrGluThrGlyGlyGluAsnAsnAspLysGluGluGluLys 400
DB 1134 GACAGAGAGACGAGGAGCTGAACTGGGGAGAGAGACACATTAAGAAAGAGAGAA 1253
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420
DB 1254 AAAGATGAGAGCTCCAGCTCTCTGAGAGCAAAATTCGCTGTCACCAACCAATTAAGATG 1313
QY 421 LysProAsnIleGluProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440
DB 1314 AAGCCAAATATTGAACCTCTGAGAAATGTGAGGTGAGGTGAGTGAACCTTCATGTTT 1373
QY 441 ArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgIleuIleGlyThr 460
DB 1374 AAGATCTCATTTGATGATCTTACTACATTAATCTTTTGCCATTCCTAGCTAATTTGGACAC 1433
QY 461 LysThrCysArgGlnValTyrGluPheArgValIleGluSerSerIleIleAlaProAla 480
DB 1434 AAAACATGTAGACAGGTGTATGAGTTAGATCAAGAGATCCAGATATCATAGACCTGTT 1493
QY 481 ProAlaGluAspValAspThrProProAlaGlyLysLysArgLysHisAlaArgIleuTyrAla 500
DB 1494 CCCACTGAGAGATGTAGACACTCTCTCCAAAGAAAGAAAGAAACATCGTGTGGGCT 1553
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520
DB 1554 GACACTGCAAGAAAGTACACTGAAAGAAAGACGGCTCTTACACATTTTTCACACTT 1613
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB 1614 CAACCTGTGACCATTCACGCGAGCTTGTGACAGTTCGTGCTGTGTGTGATGACACA 1673
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGlyCys 560
DB 1674 AATTTTGTGAAAGTTTGTCAATGTATGTACAGTGTCAAAACCGCTTCTGTGANT 1733
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrIleuAlaValArgLysCys 580
DB 1734 CGGTGCAAGACCAATGCAACCAACCAAGTGTCTTACTGCTGTCCGAGAGTGT 1793
QY 581 AspProAspLysCysLeuThrCysGlyAlaAlaAspHisIleTyrAspSerLysAsnValSer 600
DB 1794 GACCTGACCTGTGTCTGACGTGTGAGCTGTGACCAATTTGGGACAGTAAATGTATCC 1853
QY 601 CysLysAsnCysSerIleGlnAspGlySerLysLysHisLeuLeuAlaProSerAsp 620
DB 1854 TGTAAAGAACTGTAGCAATTCAGGGGCTCTTAAAGACCTTACTGTGACCGCTGTAT 1913
QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640
DB 1914 GTGGAGGCTGGGGGACTTTTATCAAGATCTCTGACAAAATAATGAATTCATTCAGAA 1973
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660
DB 1974 TACTGTGGGAGATTATTTCTCAGGATGAAGCAGACAGAAAGAAAGTGTATGACAAA 2033
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValIleAspAlaThrArgLys 680
DB 2034 TACATGTGTAGCTTTCTGTTCACATTCGAAATGATTTGTGTGTGATGACCAACCGGAAAG 2093
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700
DB 2094 GGCAACAAATTCGTTTCTTATCATTTAGATTAATCCAACTGCTATGCAAAAGTTATG 2153
QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720
DB 2154 AATGTTAATGTGTGACCAAGATAGGCAATCTTTGTGTAAGAGGCTATCCAGTGTGTGA 2213

```

QY	721	GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu	740
DB	2214	GAGTGTCTTTTGTATTCAGATACAGCCAGGCTGATGCCCTGACATATATGTGGCATTCGA	2273
QY	741	ArgGluMetGluIlePro	746
DB	2274	CGAGAAATGGAATCCCT	2291

**RESULT 13**  
 BC016391  
 LOCUS  
 DEFINITION  
 Mus musculus enhancer of zeste homolog 2 (Prosopha), mRNA (CDNA  
 clone MGC:5970 IMAGE:3586689), complete cds.  
 BC016391.1 GI:16741076  
 MGC.

**ACCESSION**  
 BC016391  
 BC016391.1 GI:16741076  
 MGC.

**VERSION**  
 BC016391.1 GI:16741076  
 MGC.

**KEYWORDS**  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

**ORGANISM**  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

**REFERENCE**  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.M., Hong, L.,  
 Diatchenko, L., Marusik, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McSwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Wortley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Heltom, B., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bonfield, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywicki, M.I., Skalka, U., Smalins, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

**JOURNAL**  
 12479932  
 2 (bases 1 to 2595)  
 Director MGC Project.

**REFERENCE**  
 Submitted (31-OCT-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

**TITLE**  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H.,  
 Kovis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navevati,  
 A.N., Gibbs, R.A.

**REMARK**  
 COMMENT  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LINT at: <http://image.llnl.gov>  
 Series: IRAC Plate: 10 Row: 1 Column: 8  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6679720  
 This clone grew slowly and was rescued by PCR.

**FEATURES**  
 Location/Qualifiers

```
source
1. .2595
/oranism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="MGC:5970 IMAGE:3586689"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(cloneal) outgrowth infected with the virus MMTV."
/clone_id="NCI CGAP_Lu29"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. .2595
/gene="B2h2"
/notes="synonym: Bux-1"
/db_xref="GeneID:14056"
/db_xref="MGI:107940"
67. .2307
/gene="B2h2"
/codon_start=1
/product="enhancer of zeste homolog 2"
/protein_id="AAH16391.1"
/db_xref="GI:1674107"
/db_xref="GeneID:14056"
/db_xref="MGI:107940"
/translacion="MGOTGKSEKGPVCMRRKRVSEYRLQLRFRRADEVKTMFSS
NRKILERTETLANOEKORRIOPVIMTSVSLGTRSCVTSIDLDPFPAQVYPIKTLN
AVASPIWMSPLQOQPMVEDVLAHNI PYMGDEVDQGFLEELIKNDGKVGHD
RECPINDEI FVELVNLGOYDDDDDDGDDPDBREKODLNDNDGKCPKPKR
PADKIFALISMPDPKTAELKELKELTQQLPGALPPTCTDGNPKAKSVORRO
SLHSFHLFCRCFKTIDFLHFAHTEVTKKNTETLNDKPGPOCYQLBGEKEP
AAALTAERIKTPRRPGRRGRRLPNNSRSTPTI SVLSKQDSDREACTGTGNG
NDKEEEKEDTSSSEANSRCQPIKKKPMI EPPENWEGAASMRVVIQTYDN
FCALRLGKITCQOVYEFVRESII LPVTEVDVDFPRKKKRLMAHCKRIOL
KKDGSNNVWVOPCDHPRODSSCPVIONFCBKCCOSSSCGSPFCRCCKAOC
NTKQCPCLTARECDPDLCTCGAADHDSDVSCNKSIDRGSKTHLLAPSVAGN
GIFLDPQKNEPISYCEGELISODEAURKGVYDKTWCSFLFNLDNFVDATRKGN
KIRFANHSVNPNCYAKVMVNGDHRIGIFARRAIQTGELFPDYRYSDALKYVGLR
REMEIP"

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 2595
Score: 3999.00 Matches: 734
Percent Similarity: 99.3% Conservative: 7
Best Local Similarity: 98.4% Mismatches: 5
Query Match: 98.7% Indels: 0
DB: 6 Gaps: 0

US-10-773-302-2 (1-746) x BC016391 (1-2595)

Oy 1 MetGlyIntHnGlyLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20
Db 67 ATGGGCGAGACTGGGAAGAAATCTGAGAAAGGAGCCGTTTGTGGCGGAAGCGTGTAAA 126
Oy 21 SerGlyTrpMetArgLysArgGlnLeuLysArgPheArgValLysArgValLysSer 40
Db 127 TCAGAGTACATGAGACTGAGACAGCTCAAGAGTTCAAGAGAGCTGAGTGAAGTAAAGACT 186
Oy 41 MetPheSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 60
Db 187 ATGTTTAGTTCCTCAATCGTCAGAAATTTTGGAAAGAACTGAAACCTTAAACCAAGTGG 246
Oy 61 LysGlnArgArgGlnGlnProValHisLeuThrSerValSerSerSerSerSerSer 80
Db 247 AAGCAGCGGAGATCAAGCTCTGCAATCATGACTTCTGTAGCTCATTTGGCGGAGACT 306
Oy 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValLysProLeuLysThr 100
Db 307 AGGAGTGTTGAGTCAACAGTGACTTGGATTTTCCAGACAAGAGTCATCCCGTTAAAGACC 366
Oy 101 LeuAsnAlaValAlaSerValProLeuMetTyrSerTyrSerProLeuGlnGlnAsnHe 120
```

```
Db 367 CTGAATCAGTCCGCTCGGTGCTTAAATGACTTGTGTGCCCTTACAAGAAATTTT 426
Oy 121 MetValGlnAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLysAsp 140
Db 427 ATGTGTGAAGACGAAATCTGTCTTACATTAACATTTCTTACATGGGGGATGAAGTCTGCAT 486
Oy 141 GlnAspGlyThrPheIleGlnGlnLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160
Db 487 CAGACCGGCACTTTCATTTGAAGAACTAATATTAATATTAATGAAGGATGATGATGATGAT 546
Oy 161 ArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlyGln 180
Db 547 AGAATATGTGATTTAATTAATGATGAATTTTGTGAGTTGCTGAATGCTCTTGTGATCA 606
Oy 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGlnArgGlnGlnLysGln 200
Db 607 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
Oy 201 LysAspLeuGlnAspHisAspAspAspLysGlySerArgProAspArgLysPheProSer 220
Db 667 AAGATCTAGAGATTAATCGAGATTAAGAACTTGCCCACTCGAAATTTCTGTCT 726
Oy 221 AspAspIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGlnLeu 240
Db 727 GATTAATATTTTGAAGCCATTTCTCTCAATGTTTCAATTAAGGCAACCGAAGAACTG 786
Oy 241 LysGlnLysTyrLysGlnLeuThrGlnGlnGlnLeuProGlnValAlaLeuProGlnCys 260
Db 787 AAGGAAAAATTAAGAACTCAACGAGACAGAGCTCCAGGTGCTGTCTCTCGAATGT 846
Oy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnIleLysSer 280
Db 847 ACTCAAAACATCGATGAGCAACCAAACTGTTTGAAGGAGGCAAGCTTGCAATTC 906
Oy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300
Db 907 TTTCAATAGCTCTTCTGTGCAATGTTTAAAGTATGATGCTCTTCAATCCTTCCAT 966
Oy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320
Db 967 GCAACACCAACACATTAAGAGAGAAACACAGAAACAGCTTGGACCAACCAACCTTGT 1026
Oy 321 GlyProGlnCysTyrGlnHisLeuGlnGlyValAlaLysGlnPheAlaAlaAlaLeuThrAla 340
Db 1027 GGACCAACAGTGTTCACGACATCTGAGGAGACTTAAGAGTGTGTGCTCTTACTGCT 1086
Oy 341 GlnArgIleLysThrProProLysArgProGlyGlyLysArgArgGlyLysArgLeuProAsn 360
Db 1087 GAGCGGATTAAGACACCACTTAACGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1146
Oy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380
Db 1147 AACAGTAGAGACCCACACACCCCAACATCATGTGCTGAGTCAAGAGATACAGACAGT 1206
Oy 381 AspArgGlnAlaGlyThrGlnThrGlyGlyGlnAsnAspLysGlnGlnGlnLys 400
Db 1207 GACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1266
Oy 401 LysAspGlnThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 420
Db 1267 AAGATGAGAGCTCAAGCTCTCTGAAAGCAAAATTTCTGGTGTCAAAACCAATAAAGATG 1326
Oy 421 LysProAsnIleGlnProProGlnAsnValGlnTrpSerGlyAlaGlnAlaSerMetPhe 440
Db 1327 AAGCAATATATTGAACCTCGTGAAGATGTGAGAGTGTGTGTAAGCTTCATGATTT 1386
Oy 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460
Db 1387 AGAGTCTCATTTGATCTTACTTACATTAATCTTTGTGTGCATTTGTAGCTTAATTTGGACC 1446
Oy 461 LysThrCysArgGlnValTyrGlnPheArgValLysGlnSerSerIleIleAlaProAla 480
Db 1447 AAAACATGTAACAGGTGTATGATTTAAGTCAAGAGTCCAGTATCATAGCACTGTT 1506
```



Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleYThr 100  
Db AGGAGTGTTCAGTCAACGAGTACTTGATTTTCCAGACAAAGTCATCCCGTTAAAGACC 399  
Qy 101 LeuAsnIleValIleAspValProIleMetYrSerTrpSerProLeuGlnIleAspPhe 120  
Db CTGAATGAGTGGCTCGGTGCTTAATGATGATCTTGTCGCCCTTAACACAGAAATTT 459  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProYrMetGluYAspGluValLeuAsp 140  
Db ATGGTGGAAAGACAACTGTTTACATAACTCTTATATGGGGGATGAAGTTCTGGAT 519  
Qy 141 GlnAspGlyThrPheIleGluGluLeuIleYAsnIleYrAspGlyYrValHisGlyAsp 160  
Db CAGATGGCATTTCATTGAAGAATAAAATTAATGATGAGAAAGTCATGCTGATAC 579  
Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnIleLeuGlyGln 180  
Db AGAGAAATGTGAATTAATGAATGAATTTTGTGGAGTTGTAAATGCTTGGTCAA 639  
Qy 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluYrGluGluYrGln 200  
Db TTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 699  
Qy 201 LysAspLeuGluAspHisArgAspAspIlyGluSerArgProProArgYrPheProSer 220  
Db AAAAGATCTAGAGATTAATCGAGATGATTAAGAAATCTGCCACCTCGGAAATTTCTGCT 759  
Qy 221 AspYrIlePheGluValIleSerSerMetPheProAspYrGlyThrAlaGluGluLeu 240  
Db GATAAATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGGACCGCAGAAAGACTG 819  
Qy 241 LysGluYrYrYrGluLeuThrGluGlnGlnLeuProGluYrAlaLeuProProGluYr 260  
Db AAAAGAAAATTAAGAACTCAGCAGACAGCAGCTCCAGGTGCTCTGCTCTGAAATGT 879  
Qy 261 ThrProAsnIleAspGlyProAsnAlaYrSerValGlnArgGluGlnSerLeuHisSer 280  
Db ACTCCAAACATCGATGAGCAAAATGCCAAATCTGTTCAGAGGAGCAAAAGCTTGCAATCA 939  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLeuYrYrAspCysPheLeuHisProPheHis 300  
Db TTTCACTAGCTCTTCTGTGACGATGTTTAAGATGATGATGCTCTCTCAATCCCTTCAT 999  
Qy 301 AlaThrProAsnThrYrYrYrAspGlyAsnThrGluThrAlaLeuAspAsnIlySerProYr 320  
Db GCACACCCCAACACATTAAGAGAAACACAGAAACAGCTTTGGAACAAGCTTGT 1059  
Qy 321 GlnYrProGlnCysYrGlnHisLeuGluGluYrAlaYrGluPheAlaAlaAlaLeuThrAla 340  
Db GGACCAACAGTGTACACATCTGAGGAGGAGCTTAAGAGATTTGCTGCTCTTACTCT 1119  
Qy 341 GlnArgGlyIleYrThrProProYrAspArgProGluYrAlaArgArgArgGlyYrArgLeuProAsn 360  
Db GAGCTTATTAAGACACCACTTAACGCCACGGGGGCCCGAGAAAGGAAAGCTTCGAT 1179  
Qy 361 AsnSerSerArgProSerThrProThrIleAsnValIleGluSerLeYrAspThrAspSer 380  
Db AAACGATACAGACCCACAGACCCCACTCACTGCTGCTGAGTCAAAGATACAGACAGT 1239  
Qy 381 AspArgGluAlaGlyThrGluThrGluYrGluYrGluAsnAsnAspYrGluGluGluYr 400  
Db GACAGAGAAGCAGGAGCTGAACCTGGGGAGAGAAACAATGATTAAGAAAGAGAAAGAA 1299  
Qy 401 LysAspArgIuThrSerSerSerSerGluIleAsnSerArgCysGlnThrProIleYrMet 420  
Db AAAAGTGAAGACGTCCAGCTCTCTGAAGCAAAATTTCTGGTGTCAAAACCAATTAAGATG 1359  
Qy 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
Db AAAGCAAAATTTGAACCTCTGAGAAATGTGAGTGAAGTGGTGGTGAAGCTCCATGTTT 1419  
Qy 441 ArgValLeuIleGlyThrYrYrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460

Db 1420 AGAGTCTCTCATGTGTACTTACTACATGATTAACCTTTTGTGCCATGTGCTAGGCTTAATGGAAACC 1479  
Qy 461 LysThrCysArgGlnValYrGluPheArgValIleGluIleSerSerIleIleAlaProAla 480  
Db AAAACATGTACACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1539  
Qy 481 ProAlaGluAspValAspThrProProArgYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYr 500  
Db CCCACATGAGATGTAGACATCTCCCTCCAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1599  
Qy 501 AlaHisCysArgYrIleGlnLeuYrValYrAspGlyYrSerSerAsnHisValYrAsnYr 520  
Db GCACATGTCAGAAAGATTAACAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659  
Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db CAACCTGTGACATCAACGAGAGCTTTGTGACAGTTCCGTGCTGTGTGATGACAA 1719  
Qy 541 AsnPheCysGluYrPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
Db AATTTTGTGAAGATTTTGTGCAATGTATGATGATGATGATGATGATGATGATGATGATG 1779  
Qy 561 ArgCysIleYrAlaGlnCysAsnThrYrGluCysProCysYrIleuAlaValArgGluCys 580  
Db CGGTGCAAGCAGATCAACCAACCAAGATGTCATGCTACCTGCTGCTGCTGCTGCTGCTG 1839  
Qy 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisArgPheSerIlyAsnValSer 600  
Db GACCTTACCTCTGTCTCAAGTGTGAGCTGTGACATTCGATGGGACATGAATAATGTATCC 1899  
Qy 601 CysIleAsnYrCysSerIleGlnArgGlySerIleYrHisIleLeuLeuAlaProSerAsp 620  
Db TGTAAAGAACTGTACATTCAGCGGGGCTTAAGAAAGCACTTACGTGGGACCGCTCGAT 1959  
Qy 621 ValIleArgIleProGlyIlePheIleYrAspProValGlnYrAsnGluPheIleSerGlu 640  
Db GTGGCAGGCTGGGCACTTTATCAAGATCCCTTAAGAAAGAAAGAAAGAAAGAAAGAAAG 2019  
Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyYrValYrAspYr 660  
Db TACTGTGGGAGATTAATTTCTCAAGATGAAGACAGACAGAAAGGATATGACAAA 2079  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgYr 680  
Db TACATGTGAGCTTTCTGTTCAACTGAACATGATTTTGTGTGATGACAAACCGAAG 2139  
Qy 681 GlnYrAsnYrIleArgPheAlaAsnHisSerValAsnProAsnCysYrAlaIleYrValIleMet 700  
Db GGCAACAAAAATTCGTTTGTCTTAATCAATTCAGTAATCAAACTGTATGCAAAAGTTATG 2199  
Qy 701 MetValAsnGluAspHisArgIleGlyIlePheAlaYrArgAlaIleGlnThrGlyGlu 720  
Db ATGGTTATGTGTACACAGATAGGCACTTTGCTTAAGAGGCTATCCAGACTGTGAA 2259  
Qy 721 GlnLeuPhePheAspYrArgYrYrSerGlnAlaAspAlaLeuYrYrValIleGlu 740  
Db GAGTTGTTTTTGTATCAAGTATACAGCAGGCTGATCCCTGAAGTATGTGGGATGAA 2319  
Qy 741 ArgGluMetGluIlePro 746  
Db CGAGAAATGGAATCCCT 2337  
  
RESULT 15  
NMU52951 2653 bp mRNA linear ROD 19-JUL-1996  
LOCUS Mus musculus putative transcriptional regulator mbrx-1 mRNA,  
DEFINITION complete cds.  
ACCESSION U52951  
VERSION U52951.1 GI:1279908  
KEYWORDS vertebrate polycomb-group gene.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

**REFERENCE**  
 1 (bases 1 to 2653)  
 Hobert, O., Sures, I., Closek, T., Fuchs, M. and Ullrich, A. Isolation and developmental expression analysis of Enx-1, a novel mouse Polcomb group gene  
 Mech. Dev. 55 (2), 171-184 (1996)

**JOURNAL**  
 PUBMED  
 8861097

**REFERENCE**  
 2 (bases 1 to 2653)  
 Hobert, O.

**AUTHORS**  
 Direct Submission

**TITLE**  
 Submitted (28-MAR-1996) Department of Molecular Biology, Massachusetts General Hospital, Wellman 8, Boston, MA 02114, USA

**JOURNAL**  
 Massachussets General Hospital, Wellman 8, Boston, MA 02114, USA

**FEATURES**  
 location/Qualifiers  
 1..2653  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /feature\_type="placenta"  
 135..2575  
 /note="putative transcriptional regulator; presumably involved in regulating the transcriptional activity of chromatin; murine homolog of the Drosophila Polcomb group (Pc-G) gene Enhancer of Zeste; contains CXC domain and SET domain"  
 /codon\_start=1  
 /product="mex-1"  
 /protein\_id="AAC52655.1"  
 /db\_xref="GI:1279909"  
 /translation="MGQNGKSEKGPVCMRRVSEVRLRQLRFRRADEVKMFSSNRQKLEETLNQMKRRIQPHIMTSSLSKTRCSTSDLPFAQVTPKLTANAVAPLIMSWSPLOQNFVBEDEVLANIPLYMGDELDQDGFIELKNDGKVHLDRECEFINDEIFVELNALGOYNDDDDDDDPDEREKQDLBNRDKTCPEPRKPADKIFEAISMFPEPKTAEBLKEKYLTEQLPGLPPECTPIIDGNPKSVQREQSLHSFHTLFCRRCFYDCFLHPFATPMYTKRKTETALDNKPGCPQCYHLEAKERAAALTAERIKTPKRGRRGRRLPNNSRSPSTPIVLSKKQDSDERAGTNGENNDKREKEDTSSSSSEANSRCQPIPKKKNIPBPBNVBSGAEBSMRYVIGITYDNFCALARLIGTKTCQVRFVRESSTIAPVEDVDTPRKKRKHRLMAHCRITQLKDGSSNNVNYQPCDHROPDCSPVIAQNFCEKQCSSECONFPCRCQAOCNTKQCPCLAVARECDPLCLTICGADHDSKVSCKNSIGRSGKHLLAPSDVAGMGIFIDPQKNBFISEYCGSLISQDEDDRGRVYDKWCSFLFNANDPVVDAETKENKIRFANHSVNPNCYAKVMVNVGDHRIIGIPARAIOTGBELFPDRIYSGADLAKYVIGIREMELP"

**ORIGIN**

**Alignment Scores:**

Pred. No.:	0	Length:	2653
Score:	3993.00	Matches:	733
Percent Similarity:	99.28	Conservative:	7
Best Local Similarity:	98.34	Mismatches:	6
Query Match:	98.64	Indels:	0
DB:	6	Gaps:	0

US-10-773-302-2 (1-746) x MMU52951 (1-2653)

QY 1 MetGlyGlnThrGlyLysLysSerGlnLysGlyProValCysTrpArgLysArgValLys 20

DB 135 ATGGGCGAGACTGGGGAAGAAATCTGAGAAAGGACCGGTTTGTGGCGAAGGCTATAAA 194

QY 21 SerGlnLysMetArgLysArgLysLysArgPheArgGlnLysArgLysValLysSer 40

DB 195 TCAGAGTACATGAGCTAGACAGCTCAAGAGTTCAGAAAGCTGATGAAGTAAAGACT 254

QY 41 MetPheSerSerAspArgGlnLysLysLeuGlnLysArgLysValLysArgLysValLys 60

DB 255 ATGTTTATGTCGAATCGTCAGAAATTTTGGAAAGAACTGAACCTTAAACCAAGAGG 314

QY 61 LysGlnLysArgGlnLysLysProValHisLysLeuThrSerValSerSerLeuArgLysThr 80

DB 315 AAGCGCGGAGAGATACACCTGTCACATCATGACTCTCTGTCAGCTCATTTGGCGGAGCT 374

QY 81 ArgGlnLysSerValThrSerAspLeuSerPheProThrGlnValLysProLeuLysThr 100

DB 375 AGGAGGTTTCAATGACCAAGGAGCTGATTTTCCAGACAAGTATCCCGTTAAAGACC 434

QY 101 LeuAlaValAlaLysSerValProLysSerTrpSerProLeuGlnLysPhe 120

DB 435 CTGAATGACATGCGCTCGCTGCTTAAATGATCTTGTGGCCCTTCAACAGAAATTTT 494

QY 121 MetValGlnAspGlnThrValLeuHisAsnLysProCysMetGlnLysGlnValLeuSer 140

DB 495 ATGTGGAGAGCGAAAGCTGTTTACATACATCTCTTAAATGGGGGATGAAGTTCTGAT 554

QY 141 GlnAspGlnThrPheLysGlnLysLysLysAsnLysArgGlnLysValHisLysVal 160

DB 555 CAGATGCGACTTTTCATTTGAAAGAACTATATAAATTTATGATGAAAGTGCATGCTAC 614

QY 161 ArgGlnLysGlnLysLysLysAsnLysLysPheValGlnLysValAlaLysGln 180

DB 615 AGAGAAATGCTGATTTATTAATGATGAATTTTGTGGAGTTGTAAATGCTCTGGTCAA 674

QY 181 TyrAsnAspAspAspAspAspAspGlnLysAspProGlnLysGlnLysGln 200

DB 675 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 734

QY 201 LysAspLeuGlnAspHisArgAspAspLysGlnLysSerArgProProArgLysPhePro 220

DB 735 AAAGATCTAGAGATTAATCGAGATGATTAAGAACTTCCCACTCGGAAATTTCTGCT 794

QY 221 AspLysLysPheGlnLysLysSerSerPheProAspLysGlnLysGlnLysGln 240

DB 795 GATAAATAATTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGCGACCGCAGAAAGACTG 854

QY 241 LysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 260

DB 855 AAAAGAAAATTTAAAGAACTACAGAGAGAGAGCTCCAGAGTCTGCTCGAAGT 914

QY 261 ThrProAsnLysAspGlnProAsnLysSerValGlnArgGlnLysSerLysSer 280

DB 915 ACTCGAAGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 974

QY 281 PheHisThrLeuPheCysArgArgCysPheLysGlnLysPheLysHisProPheHis 300

DB 975 TTTCAATGAGCTCTCTGTCGAGATGATTTTAAGTATGATGATGATGATGATGATGATGAT 1034

QY 301 AlaThrProAsnThrLysLysArgLysLysLysLysLysLysLysLysLysLysLys 320

DB 1035 GCAACACCCACATATATTAAG 1094

QY 321 GlnProGlnCysLysGlnHisLysGlnLysGlnLysLysLysLysLysLysLysLysLys 340

DB 1095 GAGACAGAGTTTACAGCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1154

QY 341 GlnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 360

DB 1155 GAGCGTAAAGAGACACCACTAAAGCGCCAGGGGGGCGCAGAGAGAGAGAGAGAGAGAG 1214

QY 361 AsnSerSerArgProSerThrProThrLysAsnValLeuGlnLysLysLysLysLysLys 380

DB 1215 AACAGTGCAGAACCCAGAGACCCCAACATCTGAGTCTGAGTCAAAAGATACAGACGT 1274

QY 381 AspArgGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 400

DB 1275 GACAG 1334

QY 401 LysAspGlnThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 420

DB 1335 AAAGATGAGAGCTCAAGCTCTCTGAAGCAATCTCGGTCTCAACCAATTAAGATG 1394

QY 421 LysProAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 440

DB 1395 AAGCAATATATTAAGCTCTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454

QY 441 ArgValLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 460

```
Db 1455 AGAGTCTCATGTGACTTACTACGATTAACCTTTGTGTCATTGCTAGGCTAATTGGAC 1514
QY 461 LySThrCysArgGlnValIYrGluPheArgValIySGIuSerSerIleIleAlaProAla 480
Db 1515 AAAACATGTATACACAGGTGTATGAGTTAGAGTCAAGAGGTCCAGATATCATAGCACTGTT 1574
QY 481 ProAlaGluAAspValAAspThrProProArgIySlySlyValArgIyVHIsArgLeuITPAla 500
Db 1575 CCCACTGAGATGTATGACACTCTCCCAAGAAAGAAAGAAACATCGGTGTGGGCT 1634
QY 501 AlaHisCysArgIySlyIleGlnLeuIySlyAAspGlySerSerAAsnHisValIYrAAsnTYr 520
Db 1635 GCACACTGCAGAAAGATACAACTGAAAAAGACGGCTCTCTAACCATGTTTACACTAT 1694
QY 521 GlnProCysAAspHisAProArgGlnProCysAAspSerSerCysProCysValIleAlaGln 540
Db 1695 CAACCTGTGACATCCACGGCAGCTGTGTGACAGTTGTCGCCCTGTGTGATAGCAAA 1754
QY 541 AAspPheCysGluIyVpPheCysGlnCysSerSerGluCysGlnAAsnArgPheProGlyCys 560
Db 1755 AATTGTGTGAAAAATTGTTCATGTAGTTCAAGTGTCAAAACCGCTTCTCGAGTGT 1814
QY 561 ArgCysIySlyAlaGlnCysAAsnThrIyGlnCysProCysTYrIleuAlaValIArgGluCys 580
Db 1815 CGGTGCMAAGCAATGCAACCAACAGTGTCCATGCTACCTGGCTGTCCGAGAGTGT 1874
QY 581 AAspProAAspLeuCysIleuThrCysGlyValIleAlaAAspHisITPAspSerIyAAsnValSer 600
Db 1875 GACCTGACCTCTGTCTACCGTGTGAGCTGTGACCATGTGGACAGTAAAAATGTATCC 1934
QY 601 CysIyAAsnCysSerIleGlnArgGlySerIySlyVHIsleuLeuLeuAlaProSerAAsp 620
Db 1935 TGTAAAGAACTGTAGCATTCACGGGGCTCTAAAAAGCACTTACTGCTGGCACCGCTGTAT 1994
QY 621 ValIleGlyITPGLyIlePheIleIyAAspProValGlnIyAAsnGluPheIleSerGlu 640
Db 1995 GTGGAGGCTGTGGCATCTTTATCAAGATCTGTACAGAAATGAAATCAATCTCAGAA 2054
QY 641 TYrCysGlyGluIleIleSerGlnAAspGluIleAAspArgArgGlyIySlyValIYrAAspIyS 660
Db 2055 TACTGTGGGAGATTAATTCTCAGGATGAGATGACAGAGGAAAGTGTATGATAA 2114
QY 661 TYrMetCysSerPheIleuPheAAsnLeuAAsnAAspPheValIleAAspAlaIThrArgIyS 680
Db 2115 TACATGTGCAGCTTCTGTCTCAACTTGAAACATGATTTTGTGTGATGCAACCCGAAAG 2174
QY 681 GlnAAsnIySlyIleArgPheAlaAAsnHisSerValAAspProAAsnCysTYrAlaIySValMet 700
Db 2175 GGCAACAAATTCGTTTGTCTATATCAGTAAATCCAAACTGTCTATGCANAAAGTTATG 2234
QY 701 MetValAAsnGlyAAspHisArgIleGlyIlePheAlaIySArgAlaIleGlnThrGlyGlu 720
Db 2235 ATGGTTATGTGACCAACAGGATAGGATCTTTGTAAAGGGCTATCCAGACTGTGAA 2294
QY 721 GlnIleuPhePheAAspTYrArgTYrSerGlnAlaAAspAlaIleuIyTYrValGlyIleGlu 740
Db 2295 GAGTTGTTTGTGATTAAGATACAGACAGGCTGATGCCCTGAAGTATGTGGGCATTCGAA 2354
QY 741 ArgGluMetGluIlePro 746
Db 2355 CGAGAAATGAAATCCCT 2372
```

Search completed: August 13, 2006, 19:50:38  
Job time : 11250.4 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 13, 2006, 12:26:26 ; Search time 1159.01 Seconds

(without alignments)  
6731.558 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGCTCKSKBKPVCWKRKRVK.....RYSQADAKYVIGIEREMRIP 746

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 348612431 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODSL=frame+ p2n.model -DEV=xlh  
-O=/abses/ABSSWEB spool/US10773302/runat\_11082006\_140221\_16507/app\_query.fasta-1  
-DB=N Geneseq -QFMT=fastap -SUFPR=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=humand0.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODSL=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=700 -MINLEN=0 -MAXLEN=200000000 -HOST=ds804  
-USER=US10773302 @CGN\_1\_1\_1518 @runat\_11082006\_140221\_16507 -NCPU=6 -ICPU=3  
-NO MMAP -NRG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_8:\*

1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*  
14: geneseqn2005s:\*  
15: geneseqn2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4051	100.0	2600	2	AAT43624
2	4051	100.0	2600	6	ABL65847
3	4051	100.0	2600	13	ADT90243

4	4051	100.0	2600	13	ADT89975	Adt89975 Human gen
5	4051	100.0	2600	13	ADT90094	Adt90094 Human gen
6	4051	100.0	2600	13	ADT90198	Adt90198 Human gen
7	4051	100.0	2711	12	ADQ22638	Adq22638 Human sof
8	4038	99.7	2558	10	ADG10641	Adg10641 Human STA
9	4038	99.7	2576	8	ABX76182	Abx76182 Lung canc
10	4038	99.7	2576	10	ADK66988	Adk66988 Gene #78
11	4038	99.7	2576	11	ADN93391	Adn93391 Cancer/an
12	4038	99.7	2576	13	ADR25182	Adr25182 Breast ca
13	4038	99.7	2576	14	ADY61813	Ady61813 Human gen
14	4030.5	99.5	2253	12	ADP95921	Adp95921 Human E2H
15	4030.5	99.5	2695	10	ADP95922	Adp95922 Human E2H
16	4024	99.3	2512	10	ADD18523	Add18523 Human pro
17	3812.5	94.1	2476	10	ADB63038	Adb63038 Human CDN
18	3106	76.7	1960	10	ADG10763	Adg10763 Human STA
19	2606.5	64.3	4185	14	ADZ61878	Adz61878 Murine Bz
20	2603.5	64.3	4606	14	ABD17754	Abd17754 Fibrinolytic
21	2603.5	64.3	4640	14	ADX05847	Adx05847 Cyclin-de
22	2601.5	64.2	2594	10	ADK66987	Adk66987 Gene #77
23	2429.5	60.0	2522	8	ABX17535	Abx17535 Human CDN
24	2156.5	53.2	2448	4	ABL11221	Ab111221 Drosophila
25	1932.5	47.7	5289	4	ABL11220	Ab111220 Drosophila
26	1202	29.7	676	13	ADQ49105	Adq49105 Novel can
27	1138	28.1	714	6	ABL93013	Ab193013 Rat metas
28	1138	28.1	714	12	ADN07804	Adn07804 Human mam
29	875	21.6	22798	4	AAK89195	Aak89195 Human dig
30	875	21.6	22798	4	AAK81124	Aak81124 Human hum
31	873.5	21.6	1108	14	ADY65573	Ady65573 S. mansoni
32	835	20.6	3162	13	ADT16786	Adt16786 Plant CDN
33	828.5	20.5	2857	13	ADX30124	Adx30124 Plant ful
34	813.5	20.1	3030	6	AAK96824	Aak96824 CDNA of M
35	809.5	20.0	3061	12	ADM47813	Adm47813 Polynucle
36	797	19.7	3173	13	ADT16470	Adt16470 Plant CDN
37	785	19.4	3148	3	AAZ34943	Aaz34943 Corn poly
38	783	19.3	3180	6	AAK96823	Aak96823 CDNA of M
39	776	19.2	485	9	ACH34755	Ach34755 Human end
40	731.5	18.1	2136	3	ADK37895	Adk37895 Fertilitea
41	731	18.0	1948	13	ADK61599	Adk61599 Plant ful
42	731	18.0	2315	2	AAZ23136	Aaz23136 Arabidops
43	731	18.0	2315	4	AAK90504	Aak90504 Arabidops
44	731	18.0	2799	8	ABZ68392	Abz68392 Nucleotid
45	722	17.8	2327	2	AAZ23137	Aaz23137 Arabidops

#### ALIGNMENTS

RESULT 1	ID	AAAT43624 standard; DNA; 2600 BP.	ALIGNMENTS
XX	AAAT43624;		
AC	25-MAR-2003 (revised)		
XX	DT 05-MAY-1997 (first entry)		
DT	05-MAY-1997 (first entry)		
XX	Chromatin regulator protein E2H2.		
XX	Chromatin; regulator; E2H1; E2H2; SUV39H; SET domain; gene therapy;		
KW	cancer; ss.		
XX	Homo sapiens.		
OS			
XX	Key	Location/Qualifiers	
FT	CDS	90..2330	
FT		/*tag= a	
XX			
PV	MO9635784-A2.		
XX	PD 14-NOV-1996.		
XX	PF 02-MAY-1996;	96WO-EP001818.	
XX	PR 10-MAY-1995;	95DE-01016776.	



XX (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
PA Jenuwein T, Laible G;  
XX  
XX WPI: 1996-518672/51.  
DR P-PSDB; AAM05260.  
XX  
XX New DNA encoding chromatin regulator protein with SET domain - and  
PT related vectors, transformed cells, proteins and antibodies, for  
PT diagnosis and treatment of cancer.  
XX  
XX Claim 1; Fig 6; 38pp; German.  
XX  
XX The DNA was isolated by screening a human B cell cDNA library with mixed  
CC Drosophila DNA probes based on the conserved SET domains in B(x) and  
CC Su(var)3-9. The DNA, and its products, are useful in therapy (esp. gene  
CC therapy) and diagnosis of human diseases that involve deregulated  
CC chromatin-regulator genes having a SET domain, esp. cancer. (Updated on  
CC 25-MAR-2003 to correct PR field.)  
XX  
XX

SO Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	7.79e-302	Length:	2600
Scores:	4051.00	Matches:	746
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	2	Gaps:	0

US-10-773-302-2 (1-746) x AAT43624 (1-2600)

QY 1 MetGlyGlnThrGlyLeuValSerGluValProValCysTrpArgValValys 20  
DB 90 ATGGGCCAGACTGGGAAAGAAATCTGAGAGGAGCAAGTTGTTGGCGAAGCGTTAAAA 149  
QY 21 SerGluTrpMetArgLeuArgGlnLeuLysArgPheArgArgAlaSerGluValLysSer 40  
DB 150 TCAGAGTACATGCGACGTGAGCAGCTCAAGAGTTCAAGCAGCTGATGAGTAAAGAGT 209  
QY 41 MetPheSerSerSerSerSerGlnLysIleLeuGluArgThrGluIleLeuArgGlnGluTrp 60  
DB 210 ATGTTTACCTCCAACTCGTCAAGAAATTTTGAAAGAAAGCAATCTTAAACCAAGAAATGG 269  
QY 61 LysGlnArgArgGlnLeuProValHisIleLeuThrSerValSerSerLeuArgGluYthr 80  
DB 270 AAACGCGAAGGATACAGCTGTGTGCAATCTGACTCTGTGAGCTCATTTGCGGAGACT 329  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 330 AGGAGGTGTTGGGTGACCAAGTGAATTTTCCACACAAAGTATCCCATTTAAAGACT 389  
QY 101 LeuAlaAlaValAlaSerValProIleMetIleTrpSerTrpSerProLeuGlnGlnAspHe 120  
DB 390 CTGAATGAGTGTGCTTCAGTACCCATTAATGTAATTTCTTGCTCTCCCTTACACAGAAATTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProIleMetGluAspGluValLeuAsp 140  
DB 450 ATGGTGAAGAGTAAAGTGTATTACATACATCTCTTAATATGGAGATGAAGTTTACAT 509  
QY 141 GlnAspGluYthrPheIleGluGluLeuLeuLysAsnIleYAspGluYValHisGluYAsp 160  
DB 510 CAGAGTGTATCTTATTTGATGAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 569  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluYgln 180  
DB 570 AGAGATGTGGGTATTAATTAATTAATTAATTTTGTGGAGTGTGTAAATGCCCTTGTGTA 629  
QY 181 TyrAsnAspAspAspAspAspAspAspGluYAspAspProGluGluArgGluGluYgln 200  
DB 630 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689

QY 201 LysAspLeuGluAspHisAspGAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACTCGGAAATTTCTTCT 749  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGluYThrAlaGluGluLeu 240  
DB 750 GATTAATAATTTTAAAGCCATTTCTCTCAATTTTCCAGATTAAGGACACAGAAAGAACTA 809  
QY 241 LysGluLysTrpYrLysGluLeuThrGluGlnGlnLeuProGluValAlaLeuProGluCys 260  
DB 810 AAGGAAAAATTAAGAACTCACCGAACACAGACTCCAGAGCCGACTTCTCTCGAATGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 870 ACCCCCAACATAGTAGCACAAAATGCTMAATCTGTTCAAGAGAGCAAGCACTTACCTCC 929  
QY 281 PheHisThrPhePheCysArgArgPheLeuYTrpCysPheLeuHisProPheHis 300  
DB 930 TTTCATACGCTTTTCTGTAGGCGATGTTTAATATGACTGCTTCCATCATCTTTTCAT 989  
QY 301 AlaThrProAsnThrYrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 990 GCAACACCAACCTTATTAAGCGAAGACACAGAAACAGCTCTAGCAACAACCTTGT 1049  
QY 321 GlyProGlnCysTrpGlnHisLeuGluValAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1050 GGACCAAGTGTATCCAGCAATTTGGAGAGCAAAAGAGTTGCTGCTCTCACCGCT 1109  
QY 341 GluArgGluLeuSerThrProProLysArgProGluYgluYArgArgArgLysArgLeuProAsn 360  
DB 1110 GAGCGATTAAGACCCCAACCAAACTCCAGAGGCCGCGAAGAGACGCTTCCCAT 1169  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1170 AAGCATGACAGGCCCAACACCCCACTTAATGTGTGAATCAAGAGTATACAGACGT 1229  
QY 381 AspArgGluAlaGluYthrGluThrGluYglnGluAsnAspLysGluGluGluGluLys 400  
DB 1230 GATGAGGAACAGGAGCTGAACCGGGGAGAGAAATATTAAGAAAGAAAGAAAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1290 AAGATGAACTTGAGACTCTCTGAGAGCAAAATTTCTGGGTGCAACACCAATTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGluAlaGluAlaSerMetPhe 440  
DB 1350 AAGCAATATATGAATCTCTGAGAAATGTGAAGTGAAGTGTGTGAAGCTCATGTGTT 1409  
QY 441 ArgValLeuIleGluYthrTrpYrAspAsnPheCysAlaIleAlaArgLeuIleGluYthr 460  
DB 1410 AGAGTCTCTATGACCTTACTATGACAAATTTCTGTGCCATGTGTAGTTAATTTGGAGCC 1469  
QY 461 LysThrCysArgGlnValYrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1470 AAACATGTAGACAGGTGTATGAGTTTAGAGTCAAAATCTTAGACATACAGCTCCAGCT 1529  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisAspLeuTrpAla 500  
DB 1530 CCGCTGAGATGTGATATCTCTCAAGAGAAAGAAAGAGAAACACCGGTGTGGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGluYSerSerSerSerSerValIleYrAsnTrp 520  
DB 1590 GCACACTGCGAAGAAATAGCTGAAGAAAGACGCTCTCTTAACATGTTTACACATAT 1649  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1650 CAACCTGTGATCATCCACGCGACCTTGTGACAGCTTGTGCTGTGTGTATGACACA 1709  
QY 541 AsnPheCysGluYrPheCysGlnCysSerSerGluCysGlnAsnAspPheProGluCys 560  
DB 1710 AATTTTGTGAAGAAATTTTGTCAATGTATGATTAAGAGTCAAAACCGCTTCCGGAGTGC 1769  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpYrLeuAlaValArgGluCys 580

```

Db      1770  CCGTCAAGGACATGCAACCAAGAGTCCCGTGTACTGCTCCGAGAGTCT 1829
Qy      581  AAPPProaPLeuCYseuThrCYsgLYaLaAaPthiSTRaPSeRiYsaYanValSer 600
Db      1830  GACCTGACCTCTGTCTTACTTGTGGAGCGCTGACCAATTTGGAGACAGTAAATGTCCTC 1889
Qy      601  CYslysaPNCysSerTllegInArgGlySerIlyseVhIleuLeuLnaLaProSerP 620
Db      1890  TCGAAGAACTGCAATTCAGCGGAGCTCCAAAAGCAATCTATGCTGACACATCTGAC 1949
Qy      621  ValaLaGlyTTPGlyYlIephelIleYsaPProValGlnLYsaNGlYnphelIleSerGlu 640
Db      1950  GTGGAGGCTGGGGGATTTTTCAAAACATCTGTGCGCAAAAAGCAATTCATCTCAGAA 2009
Qy      641  TYrCYsgLYgLYuLIleSerGlnaPGLuaLaPaRqArgGlyLYsaValTYrAspLYs 660
Db      2010  TACTGTGAGAGATATTCTCAAGATGAAGCTGACAGAAAGGAGAAAGTGTATGATAA 2069
Qy      661  TYrMetCYseRPhaLeuPheAsnLeuAsnAsnAppheValValAspAlaThrArgLYs 680
Db      2070  TACATGTCAAGCTTCTGTTCACCTTGAACAATGATTTGTGTGATGCAACCCGCAAG 2129
Qy      681  GLYAsnLYsIleArgPheAlaAsnHisSerValaAsPProAsnCYseTYrAlaLYsaValMet 700
Db      2130  GGTAAACAAATTCGTTTGCANATCATTCGGTAAATCCAACTGCTAAGCAAAAGTTATG 2189
Qy      701  MetValaNGIYsaPthiSaRgIleGlyIlePheAlaLYsaRgAlaIleGlnThrGlyGlu 720
Db      2190  ATGGTTAAGCGTGAACAGATAGTATTTTGTGCAAGAGAGCCATCCAGACTGGCGAA 2249
Qy      721  GluLeuPhePheAsPThiArgTYrSerGlnAlaAspAlaLeuLYeTYrValGlyIleGlu 740
Db      2250  GAGCTGTTTTTGTATTACAGATACAGCCAGCTGATGCCCTGAAGATGTCGGCATCGAA 2309
Qy      741  ARgGluMetGluIlePro 746
Db      2310  AGAGAAATGAAATCCCT 2327

RESULT 2
ABL65847
ID ABL65847 standard, DNA; 2600 BP.
XX
AC ABL65847;
XX
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4184.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; anti-neoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0231133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.

```

```

PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.

(AVAL-) AVALON PHARM.
PA
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI: 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 4184; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.: 7.79e-302 Length: 2600
Score: 4051.00 Matches: 746
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

```

US-10-773-302-2 (1-746) x ABL65847 (1-2600)

QY 1 MetGlyLThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB ATGGGCGACGCTGGAGAAATCTGAAGAGGACCACTTGTTGGCCGAGCGTGTAAA 149  
QY 21 SerGluTyrMetArgLysArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB TCAGAGTACATGCACCTGAGCACTCAAGAGCTTCAGACGAGCTGATGAAGTAAAGT 209  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
DB ATGTTTGGTCCATCGTCAGAAAATTTTGGAAAGAACGGAAATCTTAAACCAAGAAATGG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB AAACAGAGAGAGATACAGCTGTGCACATCTGACCTCTGTAGCTCATTCGCGGAGACT 329  
QY 81 ArgGluCysSerValThrSerAspLysPheProThrGlnValIleProLeuLysThr 100  
DB AGGAGAGTTCGAGACCAAGTGACTTGGAATTTCCACACAAAGTCATCCCATTAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB CTGAATGCAGTTGCTTCAGTACCAATATGTAATCTTGCTCCCTTACAGACAAATTTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB ATGGTGGAGATGAAACGTGTTTACATACATTCCTTATATGGAGATGAAGTTTATGAT 509  
QY 141 GlnAspGlyThrPheIleGluLysLeuLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB CAGATGGTACTTTTCATTGAAGAATATAAAATTTATGATGGGAAAGTACACGGGAGAT 569  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB AGAAGATGTGGTATTAATGATGAATTTTGTGGAGTGTGGAAAGCCCTTGCTGCA 629  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluAspGluLysGln 200  
DB TATATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 201 LysAspLeuGluAspHisAspAspAspLysGluSerArgProProArgLysPheProSer 220  
DB AAAGATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCT 749  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluLysLeu 240  
DB GATAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAGAAAGACTA 809  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGluValAlaLeuProProGluCys 260  
DB AAGGAAATATTAAGAACTCACCGAAGCAGCTCCAGGGGCACTTCTCTGTAAGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB ACCCCCAACATAGATGAGCAAAATGCTTAATCTGTTGAGAGAGACCAAAAGCTTACACTCC 929  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB TTTCATACGCTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTTCAATCCTTTTCAAT 989  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB GCAACACCCCAACACTTATAGGAGAGAAACACAGAAACAGCTCTAGACAAACCTTGT 1049  
QY 321 GlyProGlnCysTyrGlnHisLeuGluGluValLysGluPheAlaAlaLeuThrAla 340  
DB GGACACACAGTGTATCAGACATTTGAGGGGCAAAAGAGATTGCTGCTCTCACCCCT 1109  
QY 341 GluArgGlyLeuThrProProLysArgProGlyValArgArgArgGlyValArgLeuProAsn 360  
DB GAGCGAGATAAAGACCCCAACAAACGTCCAGAGGCGCGAGAAAGAGAGCGGCTTCCAAAT 1169

QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB AACAGTAGACAGGCCCAAGCACCCCAACCATTAATGTCTGGAAATCAAAAGATACAGCACT 1229  
QY 381 AspArgGluAlaGlyThrGluThrGlyValGluLysAsnAspLysGluGluGluLys 400  
DB GATAGGGAGACAGGAGCTGAACCGGGGGAGAGAAACATATATTAAGAAAGAAAGAGAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB AAGATGAAACCTTCGAGCTCTCTGAAGCAAAATTCGCGTGTCAACCAACCAATTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGluValGluAlaSerMetPhe 440  
DB AAGCAAAATATTAACCTCTGAGAAATGTGAAGTGGAGTGTGTCTGAAGCTCAATGTATT 1409  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB AGAGTCTCATGGCACTTACATGACAAATTTCTGTCCATGTCTAGTTGAATTTGGAC 1469  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB AAAACATGTACACAGGTATGATGATTAAGTCAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1529  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisAspGluTrpAla 500  
DB CCCGCTGAGGATGTGATCTCTCCAGAAAGAAAGAAAGAAAGAAACCCGGTGTGGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysValAspGlySerSerAsnHisValTyrAsnTyr 520  
DB GCACATGCGAAAGATACAGCTGAAGAAAGAGGCTCTCTTACCATGTATTAACATAT 1649  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB CAACCTGTATCATTCACCGGAGCTTGTGACAGTTGCGCCCTGTGTGATAGACAA 1709  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGlyCys 560  
DB AATTTTGTGAAAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1769  
QY 561 ArgCysLysValGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
DB CGTGCAGAGACAGTCAACACCAACAGAGTCCGCTCTTACCTGCTGCTCCAGATGT 1829  
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisTrpAspSerLysAsnValSer 600  
DB GACCTGACCTCTGTCTTACTTGTGTGAGCGGCTGACCATTTGGGACAGTAAATGTGTCC 1889  
QY 601 CysLysAsnCysSerIleGlnArgLysSerLysLysHisLeuLeuLeuAlaProSerAsp 620  
DB TGCAAGAACTGCATATATTCAGCGGGGCTCCAAAAGATCTATATGCTGGCACCACTGCAC 1949  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB GTGGCAGGCTGGGGAATTTTATCAAAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 2009  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgLysValTyrAspLys 660  
DB TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAAAGTATGATATA 2069  
QY 661 TyrMetCysSerPheLeuPheAsnLysAsnAspPheValValAspAlaThrArgLys 680  
DB TACATGTGACGCTTTCTGTCAACCTTGAACATATATTTGTGGAGATGCAACCGGCAAG 2129  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
DB GGTACCAAAATTCCTTTTGCATTAATCTGGTAAATCCAACTCTATGCCAAAGTTATG 2189  
QY 701 MetValAsnGlyAspHisAspGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
DB ATGATTAAAGGTATACAGGATGATATTTTTCCAAGAGAGCATTCAGACTGGCGAA 2249

QY 721 GluLeuPheAspTyrArgTyrSerGlnAlaAspAlaLeuIleuTyrValGlyIleGlu 740  
DB 2250 GAGCTGTTTGTGATTCAGATACAGCCGCTGATGCTCCCTGAAGTATGTCCGATCGAA 2309  
QY 741 ArgGluMetGluIlePro 746  
DB 2310 AGAGAAATCGAAATCCCT 2327

RESULT 3  
ADT90243  
ID ADT90243 standard; DNA; 2600 BP.  
XX  
AC ADT90243;  
DT 13-JAN-2005 (first entry)  
XX  
DE Human gene over-expressed in a cancerous cell, SEQ ID 356.  
XX  
KW cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
KM cytostatic; gene therapy; gene; de.  
XX  
OS Homo sapiens.  
XX  
PN WO2004091548-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 15-APR-2004; 2004WO-US009289.  
XX  
PR 15-APR-2003; 2003US-0462895P.  
XX  
PA (AVAIL-) AVAILON PHARM INC.  
XX  
PI Strovel JW, Cain CB, Horrigan SK, Augustus M;  
XX  
DR WPI; 2004-784489/77.  
XX  
PT Diagnosing cancer in a mammal comprises determining amplification of  
PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
PT mammal.  
XX  
PS Claim 3; SEQ ID NO 356; 92pp; English.  
XX  
CC The invention relates to a novel method for diagnosing cancer in a  
CC mammal. The method comprises determining amplification of specific genes,  
CC given in the specification, in the genome of a mammal. The invention  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
CC human patient; identifying an anti-neoplastic agent; determining the  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytostatic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 7,79e-302 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 13 Gaps: 0

US-10-773-302-2 (1-746) \* ADT90243 (1-2600)

QY 1 MetGlyIleThrGlyIleValSerGluIleValProValCysTrpArgIleValIleVal 20  
DB 90 ATGGCCGACACTGGGAAAGAAATCTGAGAAAGGACCAAGTTTGTGGCGAAGCTGTAAAA 149  
QY 21 SerGluTyrMetArgLeuArgGlnLeuIleValArgPheArgAlaAspGluValIleSer 40  
DB 150 TCAAGATGACATCCGACTGACAGACAGCTCAAGAGGTTCAACGACGCTGATGAAGTAAAGGT 209  
QY 41 MetPheSerSerAsnArgGlnIleValGluIleValGluIleValGluIleValGluIleVal 60  
DB 210 ATGTTAGTCCCAATCGTCAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGATCG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 270 AAACAGCAGAGAAATACAGCTGTGACATCTGACCTTCTGTAGCTCATGTGCGCGAGCT 329  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100  
DB 330 AGGAGTGTCCGTCACAGTACCTGATCTTTCATTAATGTAATCTTCCCTACAGCAGAAATTTT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnIleAsnPhe 120  
DB 390 CTGAATGCAGTTGCTTCAGTACCTACATTAATGTAATCTTGTGCTCCCTACAGCAGAAATTTT 449  
QY 121 MetValGluIleAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 450 ATGGTGAAGATGAAAGCTTTTACATACATCTTATATATGGAAGAAAGTATTTAGAT 509  
QY 141 GlnAspGlyThrPheIleGluIleValIleValAsnTyrAspGlyIleValIleValIleVal 160  
DB 510 CAGATGTGATCTTTCATTTGAAAGAACTAATTAATAATTTATGATGGAAAGTACACCGGAT 569  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 570 AGAGATGTGGTATTAATGATGAATTTTGTGGAGTTGTGTAATGCCCTTGATCAA 629  
QY 181 TyrAsnAspAspAspAspAspAspGlyIleAspAspProGluIleValGluIleValGln 200  
DB 630 TATATATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 201 LysAspLeuGluAspHisArgAspAspIleValSerAspProProArgIlePheProSer 220  
DB 690 AAAGATCTGAGGATCACCGAGATCAATAAAGAACCCGACCTCGGAAATTTCTTCT 749  
QY 221 AspIleIlePheGluAlaIleSerSerMetPheProAspIleValIleValIleValIleVal 240  
DB 750 GATAAATTTTGAAGCAATTTCTCATATGTTTCCAGATTAAGGCGACAGCAAGAAACTA 809  
QY 241 LysGluIleValIleValGluLeuThrGluGlnIleLeuProGlyAlaLeuProGluIleVal 260  
DB 810 AAGGAAATATTAAGAACTACCGAACAGACGCTCCAGCGGCACTTCTCTCAATGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaIleSerValGlnArgGluGlnSerIleHisSer 280  
DB 870 ACCCCCAATATGATGACCAATGCTAATCTGTCAGAGAGCAAACTTACACTCC 929  
QY 281 PheIleThrLeuPheCysArgArgCysPheIleValIleValIleValIleValIleVal 300  
DB 930 TTTGATACGCTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTCATCTTTTCAT 989  
QY 301 AlaThrProAsnThrTyrIleValArgIleValAsnThrGluThrAlaLeuAspAsnIleProCys 320  
DB 990 GCAACACCCCACTATATAGCGAAGAAACAGAAACGCTTATGACAAACAACTTTGT 1049  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluIleValIleValGluPheAlaAlaLeuThrAla 340  
DB 1050 GAGCACAGTGTACACGATTTTGGAGGAGCAAAAGAGTTTGTCTGCTCTCACCGCT 1109  
QY 341 GlnArgIleIleValThrProProIleValArgProGlyIleValArgArgIleValIleVal 360  
DB 1110 GAGCGGATTAAGACCCCAACCAAAAGTCCAGAGAGCGCGCAGAAAGAGACGCTTCCCAT 1169

QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1170 AACGATGACAGGCCAGCAGCCCACTATATGCTGGAATCAAGATACAGACAGT 1229  
QY 381 AspArgGluValIleGlyThrGluGlyGluAsnAspLysGluGluGluLys 400  
DB 1230 GATAGGAGAGGAGGAGTCTGAAACCGGGGAGAGAACATGATTAAGAAAGAGAGAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGluIleAsnSerArgCysGlnThrProIleLysMet 420  
DB 1290 AAAATGTAACCTTCTGAGCTCTCTGAAACAAATCTCGTCTCAACCAATTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyIleGluIleSerMetPhe 440  
DB 1350 AAGCCAAATATTGAACCTCTGAGAAATGTGAGTGTGTGCTGAAGCTCTCATGTCTT 1409  
QY 441 ArgValLeuIleGlyThrThrThrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1410 AGAGTCTCATGTGACCTTACTATGACAAATTCGTGTGCATGTCTAGGTAAATGGGACC 1469  
QY 461 LysThrCysArgGlnValIleArgIlePheArgValIleGlySerSerIleIleAlaProAla 480  
DB 1470 AAAACATGTACACAGGTATAGAGTTTGAAGTCAAGAAATCTAGCATCATAGCTCCAGCT 1529  
QY 481 ProAlaGluAspValIleAspThrProProArgLysLysLysArgLysPheIleArgLeuTrpAla 500  
DB 1530 CCCGCTGAGATGTGTGATTACTCTTCAAGAAAGAAAGAAAGAAACCGGTGTGGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValIleThrAspThr 520  
DB 1590 GCACACTGACAAAGATACACTGTAAGAAAGACGGCTCTCTTAACCAAGTTTACACATAT 1649  
QY 521 GlnProCysAspHisIleProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1650 CAACCTGTGATCATCCAGCGACGCTGTGACAGTTCCTCTGTGTGATAGCACAA 1709  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1710 AATTGTTGTGAAAGTTTGTTCATGTGTGACAGTGTCAAAACCGCTTCCGGGAGGC 1769  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValIleArgGlnCys 580  
DB 1770 CGCTGCAGAGCAAGTGTCAACCAAGCAAGTCCCGTCTACCTGCGCTGCGAGAGTGT 1829  
QY 581 AspProAspLeuCysLeuThrCysGlyValIleAspHisIleTrpAspSerLysAsnValSer 600  
DB 1830 GACCTGACCTCTGTCTTACTTGTGTGAGCCCGCTGACCTTGGACAGTAAAGATGTGTCC 1889  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAsp 620  
DB 1890 TGCAAGAACTGACAGTATTCAGCGGGGCTCCAAAAGCATCTATTGCTGGCACCATCTCAC 1949  
QY 621 ValIleArgLysPheIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1950 GTGGCAGGCTGGGGATTTTATCAAAATCTCTGTGCAAAAGAAATTCATCTTCAGAA 2009  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluIleAspArgGlyLysValIleTrpAspLys 660  
DB 2010 TACTGTGAGAGATTTATTTCTCAAGATAGAGTGCACAAAGGGAAGGTATGATTA 2069  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgLys 680  
DB 2070 TACATGTGCAGCTTCTGTCAACTTGAAACATGATTTTGTGTGAGTGCACACCGCAAG 2129  
QY 681 GlysLeuLysIleArgPheAlaAsnHisSerValAsnProAsnCysTrpAlaLysValMet 700  
DB 2130 GGTAACAAATTCGTTTGTCAAAATCATTCGGTAAATCCAAACGTGTATGCAAAAGTTATG 2189  
QY 701 MetValAsnGlyAspHisIleArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyLys 720  
DB 2190 ATGGTTAAAGGTGATCACAGATAGGTATTTTGTCCAAAGAGGCAATCCAGATCGCCAA 2249  
QY 721 GluLeuPhePheAspTrpArgTrpSerGlnIleAspAlaLeuLysTrpValGlyIleGlu 740

DB 2250 GAGCTGTTTGTGTTTACATACAGCCAGCTGATGCTCTGAAATATGTGCGCATCGAA 2309  
QY 741 ArgGluMetGluIlePro 746  
DB 2310 AGAGAAATGGAATTCCT 2327  
RESULT 4  
ADT89975  
ID ADT89975 standard; DNA; 2600 BP.  
XX  
AC ADT89975;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE Human gene over-expressed in a cancerous cell, SEQ ID 88.  
XX  
KW cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
KW cytosolic; gene therapy; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN M02004091548-A2.  
XX  
PD 28-OCT-2004.  
XX  
PP 15-APR-2004; 2004MO-US009289.  
XX  
PR 15-APR-2003; 2003US-0462895P.  
XX  
PA (AVAL-) AVALON PHARM INC.  
XX  
PI Strovel JW, Cain CB, Horrigan SK, Augustus M;  
XX  
DR WPI; 2004-764489/77.  
XX  
PT Diagnosing cancer in a mammal comprises determining amplification of  
PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
XX mammal.  
XX  
PS Claim 3; SEQ ID NO 88; 92pp; English.  
XX  
CC The invention relates to a novel method for diagnosing cancer in a  
CC mammal. The method comprises determining amplification of specific genes,  
CC given in the specification, in the genome of a mammal. The invention  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
CC human patient; identifying an anti-neoplastic agent; determining the  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytostatic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 7,79e-302 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 13 Gaps: 0  
US-10-773-302-2 (1-746) x ADT89975 (1-2600)

QY 1 MetGlyGlnThrGlyLysIleuSerGlyLysGlyProValCysThrPargLysArgValLys 20  
DB 90 ATGGGCCAAGCTGGAGAAATCTGAGAAAGGACAGCTTTGTCGGAGAGCGTGTAAAG 149  
QY 21 SerGlyUreMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
DB 150 TCAGAGTACATCGAGCTGAGACAGCTCAAGAGCTTCAGACGAGCTGATGAAGTAAAGACT 209  
QY 41 MetPheSerSerPheAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
DB 210 ATGTTTAGTTCATCGATCGTCAAGAAATTTTGAAAGAAACGAAATCTTAAACCAAGATGG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 270 AAACACCGAAGATACAGCTGTGACATCTGACTCTGTGAGCTCATTTGCGGGAGCT 329  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 330 AGGAGGTGTTGGTGCAGTGCATTTGATTTTCCAAACAAAGTCACTCCATTAAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetIYrSerTrpSerProLeuGlnLysPhe 120  
DB 390 CTGAATGCAAGTGTCTCACTACATCCAAATGATTTCTGTGCTCCCTACAGCAGAAATTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProIYrMetGlyAspGluValLeuAsp 140  
DB 450 ATGGTGGAGATGAATCGTTTACATTAATCTCTTAATGAGATGAATAGCTTTAGAT 509  
QY 141 GlnAspGlyThrPheIleGluGluLeuLysAsnIYrAspGlyLysValHisGluLysAsp 160  
DB 510 CAGATGTGATCTTCACTTGAAGAACTAAATAAATTAATGATGGAAATATACACGGGAGT 569  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluGln 180  
DB 570 AGAGATATGCGGTGTAAATGATGAAATTTTGTGGAGTTGTGATCCCTTGGTCAA 629  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluAspGluLysGln 200  
DB 630 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 690 AAAGATCTGGAGAGATCAACGAGATGATAAAGAAACCGCCCACTCGAAATTTCTTCT 749  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 750 GATTAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGAACGCAACAAACACTA 809  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGluLeuProGluValAlaLeuProGluLys 260  
DB 810 AAGGAAATATTAAGAACTCAACGAAACGAGCTCCACGCGCACTTCTCTTAATGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 870 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACACTCC 929  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysIYrAspCysPheLeuHisIleProPheHis 300  
DB 930 TTTCAATACGCTTTCTGTAGCGATGTTTAAATATGACTGCTTCTCACTCTTTCAT 989  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 990 GCAAACACCAACACTTAATTAAGGAGAAACACAAACAGCTCTGACAAACAAACCTTGT 1049  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGluAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1050 GGAACCAACAGTGTACAGATTTGAGAGGAGCAAAAGAGTTTGTCTGCTCTCAACCGCT 1109  
QY 341 GluArgIleLysThrProProLysArgProGlyLysArgArgArgGlyLysArgLeuProAsn 360  
DB 1110 GAGCGGATTAAGACCCCAACAAACGTCACAGAGGCGCGAGAAAGAGAGCGGCTTCCAT 1169  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380

DB 1170 AACATGACAGGCCCAAGACACCCCAACATTAATGTGCTGATCAAAAGATACAGACACT 1229  
QY 381 AspArgGluAlaGlyThrGluThrGlyGluAsnAsnAspLysGluGluGluLys 400  
DB 1230 GATAGGAAAGCAGGAGCTGAACCGGGGGAAGAACTAATGATTAAGAAAGAAAGAGAAAG 1289  
QY 401 LysAspGluThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1290 AAAGATGAACCTTCAGAGCTCTCTAGACAAATTTCTGGGTCAAAACCAATTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1350 AAGCAATATTTGAACCTCTGAGAAATGTGAGATGAGATGCGTCAAGAGCTCAATGTTT 1409  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaAlaLysArgLeuIleGlyThr 460  
DB 1410 AGAGTCTCAATGGCACTTAATGACAAATTTCTGTGCCAATGCTAGGTTAATGGGACC 1469  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1470 AAACATGTAGACAGGTGTATGAGTTAGAGTCAAAAGATCTAGCATACATACACTCAGCT 1529  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
DB 1530 CCGCTGAGATGTGATATCTCTCCAGAAAGAAAGAAAGAAACACCGGTTGTGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1590 GCACACTGACAAAGATACAGTGAATAAGAAAGACCGCTCTTCAACATGTTTACAACTAT 1649  
QY 521 GlnProCysAspHisPheArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1650 CAACCTGTGATCAATCCACGAGCGCTGTGACAGTTGTGCTGTGTGTATACACAA 1709  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1710 AATTTTGTGAAGTTTGTGATGTATGATGTATGATGTATGATGTATGATGTATGATGTAT 1769  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
DB 1770 CGCTGCAAAAGCAGGTGCAACCAACGAGCGCTGCTCACTGCTGCTGCTGCTGCTGCTGCT 1829  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIlePheAspSerLysAsnValSer 600  
DB 1830 GACCTGACCTCTGTCTTAATGTGTGAGCCCTGACCAATGGGACATTAATATGTCTC 1889  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleLeuLeuAlaProSerAsp 620  
DB 1890 TGCAGAACTGATGATTAACGCGGGGCTCCAAAAGCATCTAATGTGCTGACACTGATAC 1949  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1950 GTGGAGGCTGGGGGATTTTATCAAGATCTCTGTGCAAGAAATTAATGATTCATCTCAAA 2009  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAACCTGACAAAGAGGAAAGTGTATGATATAA 2069  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgLys 680  
DB 2070 TACATGTGACACTTCTGTTCATCTTAACCTTAACATGATTTTGTGGAGATGCAACCGCAAG 2129  
QY 681 GlyAsnLysIleAspPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
DB 2130 GGTAAACAAATTTGTTTGGTAAATCAATTCGTTAATCCAACTGTATGCAAAAGTTATG 2189  
QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
DB 2190 ATGTTAATCGGTGTATCAACAGATGATGATATTTTTCCAAGAGGCACTCAACATGCGCAAA 2249  
QY 721 GluLeuPhePheArgTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740

Dd	2250	GAGCTGTTTTTGTATTACAGATACGCCAGCGCTGATGCCCTGAAGTAATGTGGCATCGAA	23057
Qy	741	ArgJuncetguiljePro 746	
Dd	2310	AGAGAAATGGAATCCCT 2327	
RESULT 5			
ADT90094			
ID	ADT90094	standard; DNA; 2600 BP.	
XX	ADT90094;		
AC			
XX			
XX	13-JAN-2005	(first entry)	
Dt			
DE		Human gene over-expressed in a cancerous cell, SEQ ID 207.	
XX			
KM		cancer; detection; pre-cancerous; antineoplastic; cancer therapy;	
KW		cyrostatic; gene therapy; gene; de.	
OS			
XX	Homo sapiens.		
XX	MO2004091548-A2.		
Pd			
PD	28-OCT-2004.		
XX			
PF	15-APR-2004; 2004WO-US009289.		
PR	15-APR-2003; 2003US-0462895P.		
XX			
PA	(AVAL-) AVALON PHARM INC.		
PI	Strovel JW, Cain CB, Horrigan SK, Augustus M,		
XX			
DR	WPI; 2004-784489/77.		
XX			
PT		Diagnosing cancer in a mammal comprises determining amplification of	
PT		specific genes, e.g., serine/threonine kinase 6, in the genome of a	
PT		mammal.	
XX			
PS	Claim 3; SEQ ID NO 207; 92pp; English.		
CC		The invention relates to a novel method for diagnosing cancer in a	
CC		mammal. The method comprises determining amplification of specific genes,	
CC		given in the specification, in the genome of a mammal. The invention	
CC		further comprises methods for: diagnosing cancer or a pre-cancerous	
CC		condition in a mammal; inhibiting cancer or a pre-cancerous condition in	
CC		a mammalian cell; identifying an agent having therapeutic activity in a	
CC		human patient; identifying an anti-neoplastic agent; determining the	
CC		cancerous status of a cell; identifying a compound as an anti-neoplastic	
CC		agent; treating cancer; monitoring the progress of cancer therapy in a	
CC		patient; determining the likelihood of success of cancer therapy in a	
CC		patient; producing test data with respect to the anti-neoplastic activity	
CC		of a compound; and determining the progress of a treatment for cancer in	
CC		a patient following commencement of a cancer treatment on the patient.	
CC		The methods and compositions of the invention have cyrostatic activity	
CC		and may be used in gene therapy. The method is useful in diagnosing or	
CC		treating cancer in a mammal. This polynucleotide sequence represents a	
CC		gene over-expressed in a cancerous cell, used in the novel cancer	
CC		diagnosis method of the invention.	
XX			
SQ	Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	7,79e-302	Length: 2600	
Score:	4051.00	Matches: 746	
Percent Similarity:	100.0%	Conservative: 0	
Best Local Similarity:	100.0%	Mismatches: 0	
Query Match:	100.0%	Indels: 0	
DB:	13	Gaps: 0	
Qy	UDS-10-773-302-2 (1-746) x ADT90094 (1-2600)		
	1 MetcylgIntnHgiLyLyslsBserGuIyBgIyProValCyStrPaRglYsArGVallyvs 20		

Db	90	ATGGGCCAGACTGGGAAGAAATCTGAGAAGGACCACTTTCTTGGCCGAAGCGTAAAA	149
Qy	21	SerGlnTyrMetArgLeuArgGlnLeuIysArgPheArgAlaAspGluValLysSer	40
Db	150	TCAGAGTACAGCGACTGAGACAGACTCAAGAGGTTCCAGACGACTGTGAAGTAAAGAT	209
Qy	41	MetPheSerSerAsnAlaArgGlnLysIleLeuGlnArgGlnIleLeuAsnGlnGluTrp	60
Db	210	ATGTTTAGTTCCTCAATCGTCAGAAAAATTTTGGAAAAAGAACGAAATCTTAACCAAGAAATGG	269
Qy	61	LysGlnArgArgGlnLeuProValHisGileuThrSerValSerSerLeuArgGlyThr	80
Db	270	AAACAGCGAAGATATACAGCTCTGTGCACATCTTGACTCTTGAGACTCATTTGGGGGACT	329
Qy	81	ArgGlnCysSerValI ThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr	100
Db	330	AGGAGTGTTCGGTGAACAGTGACTTGGAATTTTCAACACAGACTCATCCCTTAAAGACT	389
Qy	101	LeuAsnAlaValAlaSerValProIleMetCysSerTrpSerProLeuGlnGlnAsnPhe	120
Db	390	CTGAATGAGATGTGCTTCACTACATCCATATGATCTTGCTGCCCTTACAGCAAAATTTT	449
Qy	121	MetValGluAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp	140
Db	450	ATGTGTGAAGATGAAACGTGTTTCAATACATCTCTTATNGGAGATGAAGTTTATGAT	509
Qy	141	GlnAspGlyThrPheIleGlnGluLeuIleIysAsnTyrAspGlyLysValHisGlyAsp	160
Db	510	CAGATGTGATCTTTCATTGAAGAATATAAATAAATATGATGCGAAAGTACACGGGGAT	569
Qy	161	ArgGlnCysArgPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln	180
Db	570	AGAGAGTGTGGGTTTATTAATGAATGAATTTTGTGAGTGGTGAATGGCCCTTGATCA	629
Qy	181	TyrAsnAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnGluGln	200
Db	630	TATATATGATGATGACGATGATGATGATGATGATGATCTTGAAAGAAAGAAAGAAAGACAG	689
Qy	201	LysAspLeuGlnAspHisIAspAspAspLysGlnSerArgProProArgLysPheProSer	220
Db	690	AAAGATCTGGAGGATACACCGAGATGATAAAGAAACCGGCCACCTCGGAAATTTCTCTCT	749
Qy	221	AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGluLeu	240
Db	750	GATAAAAATTTTGAAGCCATTTCTCTCAATGTTTCCAGATAAGGGCACACGAAAGAACTA	809
Qy	241	LysGlnLysTyrLysGlnLeuThrGlnGlnGlnLeuProGlyValLeuProProGlnCys	260
Db	810	AAAGAAATATTAAGAATCTACCGAACGACGACTCCAGCGGCATCTTCTCTGAAATGT	869
Qy	261	ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer	280
Db	870	ACCCCAACATATATGACCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACACTCC	929
Qy	281	PheHisThrLeuPheCysValArgArgCysPheLysTyrAspCysPheLeuHisProPheHis	300
Db	930	TTTCAATAGCTTTTCTGTAAAGCGATGTTTAAATATGATGCTCTTCAATCTTTTCAT	989
Qy	301	AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys	320
Db	990	GCAACACCCCAACATTAATAGCGAAGAACACAGAAACAGCTTATGACAAACAACTTGT	1049
Qy	321	GlyProGlnCysTyrGlnHisLeuGlnGluValLysGlnPheAlaAlaLeuThrAla	340
Db	1050	GGACCAAGTGTATCCAGACTTTTGAGAGGACCAAAAGAGGTTTGTGCTGCTCAACGCT	1109
Qy	341	GluArgGlnLysThrProProLysArgProGlyLysArgArgGlyValArgLeuProAsn	360
Db	1110	GAGCGGAATAAAGACCCCAACAAACGTTCCAGAGAGCGCGCAGAAAGACGCGCTTCCAAAT	1169
Qy	361	AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer	380



Db 1170 AACAGTAGCAGGCCACACCCACCATTAATGTCTGGAATCAAAAGATACAGCACT 1229  
 QY 381 AAPAAGGUAAGLYThrgLUThrgLYgLUaenAenAapLygLUgLUgLUy 400  
 Db 1230 GATAGGGAAGCAGGAGCTAAACGGGGGAGAGAAACAATGAATAAAGAAAGAGAAAG 1289  
 QY 401 LyBAAPGLuThrSerSerSerSerGluAaenSerAryCyGluThrProLlelyBmet 420  
 Db 1290 AAAGATGAACCTTCAGCTCTCTGAGCAAAATCTCGGTGTCAAAACCAATTAAGATG 1349  
 QY 421 LySProAenLleGluProProGluAenValGluTPSerGlyAAGLUaSerMetPhe 440  
 Db 1350 AAGCCAAATATTGAACCTCTCTGAGAAATGAGATGAGAGGAGGTGCTGAACCTCAATGTTT 1409  
 QY 441 ArgValLleuLleGlyThrTYrTYrAapAenPheCyGluLleAlaArgLleuLleGlyThr 460  
 Db 1410 AAGAGCTTCATGCGCACTTACTATGCAATTTCTGTGCCATTTGATGTTAATGGAGACC 1469  
 QY 461 LySThrCyAryGluValTYrGluPheAryValLyGluSerSerLleLleAProAla 480  
 Db 1470 AAAACATGTAGACAGGTGTATAGATTAGATCAAAAGAAATTAAGCATCAATAGCTCAGCT 1529  
 QY 481 ProAlaGluAapValAapThrProProArgLySlyAryAryLleAryGLeuTPAla 500  
 Db 1530 CCGGCTGAGGAGTGTGATCTCTCTCAAGAAAGAAAGAGAAACACCGGTTGTGGCT 1589  
 QY 501 AlaHLSyAryGlyLleGluLleuLySlyAapGlySerSerAenHLSyAlaTYrAenTYr 520  
 Db 1590 GCACACTGACAGAAAGATACAGCTGAAGAAAGACGGCTCTCTTAACCATTTTAAACAATAT 1649  
 QY 521 GlnProCyAapHLSyProArgGlnProCyAapSerSerCyProCyValLleAlaGln 540  
 Db 1650 CAACCTGTGATCATCCAGGAGCGCTGTGAAGTTCGTCCCTGTGTGATAGACAA 1709  
 QY 541 AenPheCyGluLyAapPheCyGlnCySerSerSerGluCyGlnAaAryPheProGlyCyS 560  
 Db 1710 AATTTTGTGAAGAAAGTTTGTCAATGTACTTCAAGTGTCAAAACCGCTTCCGGGATCC 1769  
 QY 561 ArgCySlyValGlnCyAaenThryLyGlnCyProCyTYrLleuAlaValArgLyuCyS 580  
 Db 1770 CCTGTCAAAAGCAGGACAAACCAAGAGTCCCGCTGACTGCTGTCCGAGAGTGT 1829  
 QY 581 AAPProAapLyuCyLleuThrCySGLyAlaAaAryHLSyAryAryAryAryAryAryAry 600  
 Db 1830 GACCTGACCTGTGTACTGTGTGAGCGCTGACCATTTGAGCAAGTAAATGTGTCC 1889  
 QY 601 CySlyAaenCySerLleGlnAryGlySerLySlyHLSyLleuLleuLleAProSerAap 620  
 Db 1890 TGCAGAACTGCAGATTCAGCGGGGCTCCAAAAGCATCTATTGCTGGCACCATCTGAC 1949  
 QY 621 ValAlaGlyTPGlyLlePheLleLyAapProValGlnLyAaenGluPheLleSerGlu 640  
 Db 1950 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCGTGACAAAGAAATGAATTCATCTCAGA 2009  
 QY 641 TYrCySGLyGluLleLleSerGlnAapGluAaAryAryAryGlyValTYrAryLyS 660  
 Db 2010 TACTGTGAGAGATTAATTTCTCAAGTGAAGCTGACAGAAAGGGAAGTGAATGAATAA 2069  
 QY 661 TYrMetCySerPheLyuPheAenLyuAaenAapPheValValAapAlaThryAryLyS 680  
 Db 2070 TACATGTGACGCTTCTGTCTTCAACTTGAACATGATTTTGTGTGATGACACCGGCAAG 2129  
 QY 681 GlysAenLySLeAryPheAlaAenHLSySerValAaenProAenCyTYrAlaLySValMet 700  
 Db 2130 GGTAACAAATTCGTTTCAATCATCTCGTAATTCCAAATCGCTATCAAAAGTTAAG 2189  
 QY 701 MetValAaenGlyAaenHLSyAryGlyLlePheAlaLyAryAryAlaLleLlnThryGlu 720  
 Db 2190 ATGGTTAAAGGTGATCAGAGATAGATATTTTTCCAAAGAGCCATTCAGATCGCGCAA 2249  
 QY 721 GlnLyuPhePheAapTYrAryTYrSerGlnAlaAapAlaLleuLyTYrValGlyLleGlu 740  
 Db 2250 GAGCTGTTTTTGTATTACGATACAGCCAGGCTGATGCCCTGAAGATTCGGCATCGAA 2309

QY 741 ArgGluMetGluLlePro 746  
 Db 2310 AGAGAAATGAAATCCCT 2327  
 RESULT 6  
 ADT90198  
 ID ADT90198 standard; DNA; 2600 BP.  
 AC ADT90198;  
 XX  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE Human gene over-expressed in a cancerous cell, SEQ ID 311.  
 XX  
 KW cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
 KW cytostatic; gene therapy; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO2004091548-A2.  
 XX  
 PD 28-OCT-2004.  
 XX  
 PF 15-APR-2004; 2004MO-US009289.  
 XX  
 PR 15-APR-2003; 2003US-0462895P.  
 XX  
 PA (AVAL-) AVALON PHARM INC.  
 XX  
 PI Strovel JW, Cain CB, Horrigan SK, Augustus M;  
 XX  
 DR WPI; 2004-784489/77.  
 XX  
 PT Diagnosing cancer in a mammal comprises determining amplification of  
 PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
 PT mammal.  
 XX  
 PS Claim 3; SEQ ID NO 311; 92pp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing cancer in a  
 CC mammal. The method comprises determining amplification of specific genes,  
 CC given in the specification, in the genome of a mammal. The invention  
 CC further comprises methods for: diagnosing cancer or a pre-cancerous  
 CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
 CC a human patient; identifying an agent having therapeutic activity in a  
 CC human patient; identifying an anti-neoplastic agent; determining the  
 CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
 CC agent; treating cancer; monitoring the progress of cancer therapy in a  
 CC patient; determining the likelihood of success of cancer therapy in a  
 CC patient; producing test data with respect to the anti-neoplastic activity  
 CC of a compound; and determining the progress of a treatment for cancer in  
 CC a patient following commencement of a cancer treatment on the patient.  
 CC The methods and compositions of the invention have cytostatic activity  
 CC and may be used in gene therapy. The method is useful in diagnosing or  
 CC treating cancer in a mammal. This polynucleotide sequence represents a  
 CC gene over-expressed in a cancerous cell, used in the novel cancer  
 CC diagnosis method of the invention.  
 XX  
 SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 7.79e-302 Length: 2600  
 Score: 4051.00 Matches: 746  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 13 Gaps: 0  
 US-10-773-302-2 (1-746) x ADT90198 (1-2600)  
 QY 1 MetGlyGlnThrgLyLySlySerGluLyAapValCyBTPAryGlyAryValLyS 20



Db 90 ATGGGCGACACTGGGAGAGAAATCTGAGAGGAGCCAGTTTGTGGCGGACCGTGTAAA 149  
Qy 21 SerGluYrMetCArgLeuArgGluLeuYsArgPheArgAlaAspGluValIleYsSer 40  
Db 150 TCAGAGTCAATCCAGCTGAGACAGCTCAAGAGGTTCAACAGAGCTGATGAAGTAAAGGT 209  
Qy 41 MetPheSerSerAsnArgGluLeuYsIleLeuGluValArgThrGluIleLeuAsnGlnGluTrp 60  
Db 210 ATGTTAGTTCACATCCGTACAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATGG 269  
Qy 61 LysGlnArgArgGluLeuProValHisIleLeuThrSerValIleSerSerLeuArgGluYrThr 80  
Db 270 AAACGCGAAGATATACAGCTGTGACATCTGACTCTGTGAGGTTCACTTGGCGGAGCT 329  
Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuYsThr 100  
Db 330 AGGAGGTTCGTGTCACAGTACCTTGATTTTCCACACAAAGTCATCCATTAAAGACT 389  
Qy 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPhe 120  
Db 390 CTGAATGCAATGGCTTCAGTACCCATTAATGTAATCTGTGCTCCCTACAGCAGAAATTTT 449  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProYrMetGluAspGluValLeuAsp 140  
Db 450 ATGTGGAGAGATGAATCGTTTACATTAATCTTATATGGAGATGAAGTTTATAGAT 509  
Qy 141 GlnAspGluYrThrPheIleGluGluLeuIleLeuAsnYrAspGluYsValHisGluYsAsp 160  
Db 510 CAGAGATGTAATCTTCACTTGAAGAACTAATTAATAAATTTATGATGGGAAAGTACAGCGGAT 569  
Qy 161 ArgGluCysGluYrPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluYsGln 180  
Db 570 AGAGATGTGGGTATTAATAATGATGAATTTTGTGGAGTTGGTGAAGCCCTTGCTGATA 629  
Qy 181 TyrAsnAspAspAspAspAspAspGluYsAspAspProGluGluArgGluGluYsGln 200  
Db 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
Qy 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
Db 690 AAAAGATCGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTCTTCT 749  
Qy 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGluYrThrAlaGluGluLeu 240  
Db 750 GATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAACACTTA 809  
Qy 241 LysGluYsYrYrLysGluLeuThrGluGlnGluLeuProGluYsAlaLeuProProGluYs 260  
Db 810 AAGGAAATATTAAGAACTACCCAAACAGAGCTCCAGGCGCACTTCTCTGATATGT 869  
Qy 261 ThrProAsnIleAspGluYrProAsnAlaYsSerValGlnArgGluGlnSerLeuHisSer 280  
Db 870 ACCCCCAACATGATGAGCAAAATGCTAATCTGTTCAGAGAGACAAAGCTTACATCCC 929  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLeuYrYrAspCysPheLeuHisProPheHis 300  
Db 930 TTTCAATAGCTTTTCTGTAGGCGATGTTTAAATATGATCTGCTTCTCAATCCCTTTTCAAT 989  
Qy 301 AlaThrProAsnThrYrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
Db 990 GCAAACCCCAACATTAATTAAGCGAAGAACACAGAAACAGCTTGAACAACAAACCTTGT 1049  
Qy 321 GlnProGlnCysYrGlnHisIleLeuGluGluYsAlaYsGluPheAlaAlaIleLeuThrAla 340  
Db 1050 GGAACCAAGTGTATACAGATTTTGAAGGAGCAAAAGAGTTTGCCTGCTCTCAACCGCT 1109  
Qy 341 GlnArgGluIleYsThrProProLysArgProGluYsAlaYsArgArgGluYrArgLeuProAsn 360  
Db 1110 GAGCGGATTAAGAACCCCAACCAAACTCCAGAGAGCCCGCAGAAAGGAGCGGCTTCCCAAT 1169  
Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
Db 1170 AACGATAGAGGCGCCAGCACCCCAACATTAATATGTGGAATCAAAAGATACAGACAGT 1229

Qy 381 AspArgGluAlaGluYrThrGluThrGluYsGluAsnAsnAspLysGluGluGluYs 400  
Db 1230 GATAGGAGAACAGAGGATCTGAACACGGGGGAGAGAACATATATTAAGAAAGAAAGAAC 1289  
Qy 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleYsMet 420  
Db 1290 AATGATTAACCTTGACCTCTCTGAAACAAATTTCTGGGTGTCAACACCAATTAAGATG 1349  
Qy 421 LysProAsnIleGluProProGluAsnValGluTrpSerGluYsAlaGluAlaSerMetPhe 440  
Db 1350 AAGCAAAATATTGAACCTCTGAGAAATGTGAGAGTGAAGTGTGTGAAGCTCAATGTGT 1409  
Qy 441 ArgValLeuIleGluYrYrYrYrAspAsnPheCysAlaIleAlaAlaGluIleGluYrThr 460  
Db 1410 AGAGTCTCTCATGGCATTACTATGACAAATTTCTGTGTCATTTGCTAGATTAAATGGGACC 1469  
Qy 461 LysThrCysArgGlnValYrGluPheArgValIleGluIleSerSerIleIleAlaProAla 480  
Db 1470 AAAACATGTAGACAGGTGATGAGTTTAAAGTCAAAAGAACTTACATCATAGCTCCAGCT 1529  
Qy 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
Db 1530 CCGCTGAGAGATGTGATACTCTCCAAAGGAAAAGAAAGAAACACCGGTTGGGCT 1589  
Qy 501 AlaHisCysArgLysIleGlnLeuYsLysAspGluYsSerSerAsnHisValYrAsnYr 520  
Db 1590 GCACATGCGAAGAAAGATACAGCTGAAGAAAGACCGCTCTCTTAACCTATGTTTACACATAT 1649  
Qy 521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1650 CAACCTGTGATATCCACCGGACCTTGTGACAGTTCTGTGCTGTGTGTGATGACAA 1709  
Qy 541 AsnPheCysGluYsPheCysGlnCysSerSerGluCysGlnAsnAspPheProGluYs 560  
Db 1710 AATTTTGTGAAGATTTTGTCAATGATGATCAAGATGCAAAACCGCTTCCGGAGATGC 1769  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrYrLeuAlaValArgGluCys 580  
Db 1770 CGCTGCAAGCAGAGTCAACACCAACAGAGCCCGGTCTTACCTGTGCTCCGAGAGTGT 1829  
Qy 581 AspProAspLeuCysLeuThrCysGluYsAlaAlaAspHisThrAspSerLysAsnValSer 600  
Db 1830 GACCTGACCTCTCTCTTATCTGTGAGCCGCTCAACATTTGGACATTAATAATGTGTCC 1889  
Qy 601 CysLysAsnCysSerIleGlnArgIleYsLysLysHisIleLeuLeuAlaProSerAsp 620  
Db 1890 TGCAAAGACGCAATATTACGGCGGCTCCAAAAGCATCTATTGCTGGCACCATCTGAC 1949  
Qy 621 ValAlaGluYrProGluIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 1950 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGTGCAAGAAATGAATTCATCTCGAA 2009  
Qy 641 TyrCysGluGluIleIleSerGlnAspGluAlaAspArgArgGluYsValYrAspLys 660  
Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTACAGAAAGGAAAGTATATGATTA 2069  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValaAspAlaThrArgLys 680  
Db 2070 TACATGTGAGCACTTCTGTTCATCTTGAACATGATTTTGTGGTGAAGTGAACCGCAAG 2129  
Qy 681 GlnLysLysIleArgPheAlaAsnHisSerValAsnProAsnCysYrYrAlaLysValMet 700  
Db 2130 GGTAAACAAATTTCTTTTGGCAAAATCATTCGGTAAATCAAACTGCTATGCAAAAGTTATG 2189  
Qy 701 MetValAsnGluYsAspHisArgIleGluYrIlePheAlaYsArgAlaIleGlnThrGlu 720  
Db 2190 ATGTGTACGGGTATCACAGATAGGTATTTTTCAGAGAGGCCATCCAGACTGGGGA 2249  
Qy 721 GlnLeuPhePheAspYrArgYrSerGlnAlaAspAlaLeuLysYrYrValGluYsGlu 740  
Db 2250 GAGCTGTTTTTGTATTACAGATACAGCCAGGCTATGCCCTGAAGTATGTGGCATCGAA 2309

QY 741 ArgGluMetGluIlePro 746  
DB 2310 AGAGAAATGGAAATCCCT 2327  
RESULT 7  
ADQ22638  
ID ADQ22638 standard; DNA; 2711 BP.  
XX  
AC ADQ22638;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5458.  
XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM ds.  
XX Homo sapiens.  
OS  
XX MO2004048938-A2.  
PN  
XX 10-JUN-2004.  
PD  
XX 26-NOV-2003; 2003WO-US038193.  
PF  
XX 26-NOV-2002; 2002US-0429739P.  
PR  
XX (PROT-) PROTEIN DESIGN LABS INC.  
PA  
XX Aziz N, Ginsburg WM, Zlotnick A;  
PI  
XX WPI; 2004-441208/41.  
DR  
XX  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
XX Example 2; SEQ ID NO 5458; 210pp; English.  
PS  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2711 BP; 865 A; 547 C; 658 G; 641 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 8.23e-302 Length: 2711  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 12 Gaps: 0  
US-10-773-302-2 (1-746) x ADQ22638 (1-2711)  
QY 1 MetGlyGlnThrGlyLeuIleProValCysThrArgGlySerValIle 20  
DB 167 ATGGGCGCAGACTGGGAGAAATCTGAGAGGAGCACTTTGTGGCGGAAAGGTATAAA 226  
QY 21 SerGluTyrMetArgLeuArgGlnLeuIleValArgPheArgAlaAspGluValIle 40  
DB 227 TCAGGTCATCCGACCTGAGACAGCTCAAGAGCTTCAAGCGAGCTGATGAATAAAGACT 286

QY 41 MetPheSerSerAsnArgGlnIleLeuGluArgThrGlnIleLeuAsnGlnGluTrp 60  
DB 287 ATGTTAGTCCCATCGTCAGAAATTTTGGAAAGACGAAATCTTAAACCAAGAAAG 346  
QY 61 LysGlnArgArgIleGlnProValIleIleuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACGCGAAGAGATACAGCTCTGTGCAATCTCGATCTGTGTAGCTTATGGCGGAGCT 406  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100  
DB 407 AGGAGTGTGCGTACCGACGACGACCTGATTTTCCACACAGTCATCCATTAAGACT 466  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnIleAsnPhe 120  
DB 467 CTGAATGACAGTGTCTCGATCCCAATATGATTTTGTGCTCCCTACGAGAAATTT 526  
QY 121 MetValGluAspGluThrValLeuIleAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 527 ATGTCGAAAGATGAAACTGTTTACATTAACATCTCTTATATGGAGATGAAGTTTAGAT 586  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleIleAsnTyrAspGlyIleValIleGlyAsp 160  
DB 587 CAGATGTGACTTTCATTAAGAACTAATAAAATTAATGATGGGAAAGTACACGGGAGT 646  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 647 AGAGATATGGGTTTAAATGAATGAATATTTTGTGGAGTGGTAAAGCCCTTGTCTCA 706  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluValArgGluGln 200  
DB 707 TTTATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
QY 201 LysAspLeuGluAspPheIleArgAspAspIleGluSerArgProProArgIleAspPhePro 220  
DB 767 AAAGATCGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 826  
QY 221 AspIleIlePheGluAlaIleSerSerMetPheProAspIleGlyThrAlaGluGluLeu 240  
DB 827 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGTAAAGGCGACGCGAAGAACTCA 886  
QY 241 LysGluIleValGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGluCys 260  
DB 887 AAGGAAATTTTAAAGAACTACCGACAGAGCTCCAGGCGCACTTCTCTGAATGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaIleSerValGlnArgGluGlnSerIleIleSer 280  
DB 947 ACCCCACATATGATGACCAAAATGCTTAATCTGTTGAGAGAGCAAAAGCTTACATCC 1006  
QY 281 PheIleThrLeuPheCysArgArgCysPheIleValIleValIleIleProPheIle 300  
DB 1007 TTTCAATGCTTTTCTGTAGCGATGTTTAAATATGATCTGCTTCAATCCCTTTTAT 1066  
QY 301 AlaThrProAsnThrTyrIleArgIleAsnThrGluThrAlaLeuAspAsnIleAspProCys 320  
DB 1067 GCACACCCCACTTATTAAGCGGAAGAACACAGAAACAGCTTATGACACAAACCTTGT 1126  
QY 321 GlnProGlnCysTyrGlnIleLeuGluIleValIleIleValIleValIleValIle 340  
DB 1127 GGACACAGTGTTCACAGCACTTTGGAGGAGCAAAAGAGTTTGTGCTGCTCAACCGCT 1186  
QY 341 GlnArgIleIleValThrProProIleArgProGlyValIleValIleValIleValIle 360  
DB 1187 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGAGCCCGACAGAGAGAGCGCTTCCCAT 1246  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleGluIleAspThrAspSer 380  
DB 1247 AACGATACAGAGCCCAAGACCCCACTTATATGTCTGGAATCAAGATACAGACAGT 1306  
QY 381 AspArgIleValIleGlyThrGlyIleGlyGluAsnAspIleGluGluGluIle 400  
DB 1307 GATGCGGAAGACGAGGACTGAACCGGGGGAAGAAACATGATTAAGAAAGAAAGAGAG 1366

[illegible]

ID	ADG10641		
XX	ADG10641 standard; cDNA; 2558 BP.		
AC			
XX	ADG10641;		
DT	26-FEB-2004 (first entry)		
XX			
DE	Human STAT6-activating protein-encoding cDNA, SEQ ID NO:231.		
XX			
KM	Human; STAT6-activating protein; drug screening; activator; inhibitor; allergic diseases; inflammation; autoimmune disease; diabetes; hyperlipidaemia; cancer; infection; HIV infection; antiallergic; human immunodeficiency; cancer; Th1 hypofunction; anti-HIV; antiinflammatory; antidiabetic; antiparasitic; antineoplastic; anti- cyclostatic.; gene therapy; antisense therapy; ribozyme therapy; gene; ss		
OS	Homo sapiens.		
XX			
PN	WO200296943-A1.		
XX			
PD	05-DEC-2002.		
XX			
PF	22-MAY-2002; 2002WC-JP004949.		
XX			
PR	25-MAY-2001; 2001JP-00157043. 30-AUG-2001; 2001JP-00260681. 10-OCT-2001; 2001JP-00313175.		
XX			
PA	(ASAH ) ASAH KASEI KOGYO KK.		
PI	Honda G, Matsuda A, Muramatsu S, Iohizawa K; WPI; 2003-140442/13. P-PsDB; ADG10642.		
DR			
Pt	STAT6-activating proteins and encoded genes, applicable in diagnosis of and developing drugs to treat allergic diseases, inflammations, autoimmune diseases, diabetes, hyperlipidaemia, infections e.g. HIV, and cancer.		
PS	Claim 4; SEQ ID NO 231; 2080bp; Japanese.		
CC	The invention relates to 242 human STAT6-activating proteins and cDNAs encoding them (ADG10411-ADG10894) and to sequences with 95% or more homology to the STAT6-activating proteins and their encoding nucleic acids. The invention also relates to recombinant vectors and host cells comprising a STAT6-activating protein-encoding nucleic acid; the recombinant production of a STAT6-activating protein; an antibody specific for a STAT6-activating protein; antisense oligonucleotides and ribozymes targeted to nucleic acids encoding a STAT6-activating protein; methods of screening for activators or inhibitors of STAT6-activating proteins; drug compositions comprising a modulator of STAT6-activating protein activity or expression; and methods of treating patients by administration of the drug compositions. The STAT6-activating proteins, nucleic acids encoding them, and modulators of their activity or expression are useful in the diagnosis and treatment of allergic diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia, cancer, infections (e.g., HIV), cancer and disorders associated with Th1 hypofunction. The present sequence is related to the invention.		
SQ	Sequence 2558 BP; 808 A; 513 C; 612 G; 625 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	7,648-301	Length:	2558
Score:	4038.00	Matches:	744
Percent Similarity:	99.7%	Conservative:	0
Best Local Similarity:	99.7%	Mismatches:	2
Query Match:	99.7%	Indels:	0
DB:	10	Gaps:	0

US-10-773-302-2 (1-746) x ADG10641 (1-2558)

```

Db      58 ATGGGCGACAGCTGGGAGAAATCTGAGAGGGACCGAGTTGTTGGCGGAAGCGTGAATA 117
Qy      21 SerGIuTYMeArGLuArgLnuLeuLYaRPhEArGLaAspGIuValLYsSer 40
Db      118 TCAGAGTATCATCGCATGAGACAGCTCAAGAGGTTCAAGACGATGATGAAGTAAGAAGT 177
Qy      41 MetPhESeRSeRAspArgLnuLYsLIeUuGIuAaRgThGluLIeUuAsnGInGInuTrp 60
Db      178 ATGTTTAGTTCAGATGTCAGAAAATTTTGAAAAGACGAAATCTTAAACCAAGATG 237
Qy      61 LYSGInARgArGIleGInPRoVAlhISILeUThSeRSeRSeRLeuArgLYThR 80
Db      238 AAACGCGAAGATACAGCTGTGCACATCTGACTTCTGTGACCTCAATGGCGGAGCT 297
Qy      81 ArgGIuCYSeRSeRValThSeRAspLeuAspPheProThrGInValIleProLeuLYThR 100
Db      298 AGGAGGTGTTCCGTGACAGTACCTTGATTTCCAAACAAGTCATCCCATTAAGAAGCT 357
Qy      101 LeuAsnAlaValAsESeRValProLIeMeTYrSeRTrpSeRProLeuGInGInAsnPhE 120
Db      358 CTGAATGACGTTGCTTCAGTACCCCAATGATTTCTTGCTCCCTACAGACGAATTT 417
Qy      121 MecValGIuAspGIuThRValIeUhISaEnIleProTYrMeGIYAspGIuValIeUAsp 140
Db      418 ATGCTGGAAGATGAAGAACTTTTACATACATTCCTTAATGAGAGATGAAGTTTAGAT 477
Qy      141 GIuAspGIuThRPhEIIeGIuGInuLeuIleLYsAsnTYrAspGIuLYsValhISGLYAsp 160
Db      478 CAGCATGTGACTTTCATGAAAGACTTAATAAAATTTATGATGGAAAGTACACGGGGAT 537
Qy      161 ArgGIuCYSeGIYPhEIIeAsnAspGIuLIePhEValGIuLeuValAsnAlaLeuGIn 180
Db      538 AAGAGATGGGGTTTAATAATATGATAAATTTTGTGGAGTGGTGAATGCCCTTGCTCA 597
Qy      181 TYrAsnAspAspAspAspAspAspAspGIYAspAspProGIuGIuArgLIuLYsGIn 200
Db      598 TTAATATGATGATGACATGATGATGATGAGACGATCTCGAAGAAAGAAAGAAAGCAG 657
Qy      201 LYsAspLeuGIuAspPhISaRgAspAspLYsGIuSeRArgProProArgLYsPheProSeR 220
Db      658 AAAGATCTGAGAGATCACCGAGATGATAAAGACCGCCACCTCGGAATTTCTTCTCT 717
Qy      221 AspLYsLIePhEGLuAlaIleSeRSeRMeTYrProAspLYsGIYThRAlaGInuGInu 240
Db      718 GATTAATAATTTTGAGCGCATTTCTCAATGTTTCCAGATTAAGGACACAGCAAGAACATA 777
Qy      241 LYsGIuLYsTYrLYsGIuLeuThRGIuGInGInuLeuProGIYAlaLeuProProGIuCYs 260
Db      778 AAGGAAAAATTAAGAACTCACGAAACGACGCTCCAGCGGCACTTCTCTGAATCT 837
Qy      261 ThRProAsnLIeAspGIYProAsnAlaLYsSeRValGInArgGInuGInSeRLeuHISer 280
Db      838 ACCCCCAACATGATGAGACCAATGCTAAATCTGTTCAAGAGACCAAAAGCTTACATCC 897
Qy      281 PhehISThRLeuPhEArGLuArgRGCYsPheLYsTYrAspCYsPheLeuHISProPheHIS 300
Db      898 TTTCAATACGCTTTTCTGTAAGCGATGTTTAAATATGACTGCTTCTCATCTCTTTTAT 957
Qy      301 AlaThRProAsnThRTYrLYsArGLYsAsnThRGIuThRAlaLeuAspAsnLYsProCYs 320
Db      958 GCAACACCAACACTTAATAGCGGAAAGAACACAGAAACAGCTCTAGACAAACCAACTTGT 1017
Qy      321 GIYProGInCYeTYrGInhISLeuGInuLYsAlaLYsGIuPheAlaIleAlaLeuThRAla 340
Db      1018 GGACCAACAGTGTATCAGCATTTGAGGAGGCAAAAGAGTTGCTGCTCTCACCCCT 1077
Qy      341 GIuArgLIeLYsThRProProLYsArgProGIYGIYArgArgArgGLYArgLeuProAsn 360
Db      1078 GAGCGGATTAAGACCCCAACCAAAAGCTCCAGGAGCGCGAGAAAGGACGAGGTTCCCAT 1137
Qy      361 AsnSeRSeRArgProSeRThRProThRIleAsnValIeUuGIuSeRLYsAspThRAspSeR 380

```

---

```

Db      1138 AACAGTAGAGGCCACAGACCCCAACATTAATGTCGGAATCAAGAGTACAGACAT 1197
Qy      361 AspArgGIuAlaGILYThGInuThRGIYGIuAsnAsnAspLYsGInuGInuLYs 400
Db      1198 GATAGGAAAGCAGGACCTGAACCGGGGAGAGAAACAATGTATTAAGAAAGAAAGAAAG 1257
Qy      401 LYsAspGIuThRSeRSeRSeRSeRGIuAlaAsnSeRArgCYsGInThRProIleLYsMeC 420
Db      1258 AAGATGAACCTTCAGACTCTCTGAGACCAATTCCTGGTTCAAACACCAATTAAGATG 1317
Qy      421 LYsProAsnLIeGIuProProGIuAsnValGIuTrpSeRGIYAlaGIuAlaSeRMePhE 440
Db      1318 AAGCCAAATATTGAACCTCTGAGAAATGGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 1377
Qy      441 ArgValIeUuIleGIYThRTYrTYrAspAsnPhEcyValaIleAlaArgLeuIleGIYThR 460
Db      1378 AGAGTCCCAATGGGACCTTACTATGACATTTCTGTGTCATTTGCTAGGTTAATGGGACC 1437
Qy      461 LYsThRCYsArgGInValTYrGIuPheArgValLYsGIuSeRSeRLeIleAlaProAla 480
Db      1438 AAACATGTAGACAGTGTATGATTAAGTCAAGAAATCTAGACATCATAGCTCACCT 1497
Qy      481 ProAlaGIuAspValAspThRProProArgLYsLYsLYsArgLYsHISaRGLeUThRAla 500
Db      1498 CCGGCTGAGATGTGAGATACCTCTCCAAAGAAAAGAAAGAAAGAACACCGGTTGGGCT 1557
Qy      501 AlhISCYsArgLYsLIeGInuLeuLYsAspGIYSeRSeRAsnHISValTYrAsnTYr 520
Db      1558 GCACACTGCAGAAAGATCACCTGAAAAAGACCGCTCTCAACATGTTTACAACTAT 1617
Qy      521 GInProCYsAspPhISaRgArgGInProCYsAspSeRSeRProCYsValIleAlaGIn 540
Db      1618 CAACCTGTGATCATTCACGGGAGCTTGTAAGTGTGACCTTGATGTATGACADA 1677
Qy      541 AsnPhEcyGIuLYsPhEcyGInCYsSeRSeRGIuCYsGInAsnArgPheProGIYCYs 560
Db      1678 AATTTTGTGAAAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAACCGCTTCCGGAGGC 1737
Qy      561 ArgCYsLYsAlaGInCYsAsnThRLYsGInCYsProCYSTYrLeuAlaValArgLIuCYs 580
Db      1738 CGCTGCAAAAGCAGTGCACACCAAGACAGTGCCTGCTCACTGCTGCTGCGAGAGTGT 1797
Qy      581 AspProAspLeuCYsLeuThRCYsGIYAlaAlaAspPhISThRAspSeRLYsAsnValSeR 600
Db      1798 GACCTGACCTGTGTCTTACTTGAGACCTGTACATGTGGACAGTAAAGTGTGCTC 1857
Qy      601 CYsLYsAsnCYsSeRLIeGInArgLYSeRLYsLYsHISLeuLeuAlaProSeRAsp 620
Db      1858 TGCAAGAACTGCAGATTAACAGCGGGGCTCCAAAAGCATCTATTGTGGACCATCTGAC 1917
Qy      621 ValAlaGILYThRGIYLIePhEIIeLYsAspProValGInLYsAsnGInuPheIISerGIu 640
Db      1918 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAGAAAATTAATCATCTCAGAA 1977
Qy      641 TYrCYsGIYGIuLIeIleSeRGIuAspGIuAlaAspArgArgGIYLYsValTYrAspLYs 660
Db      1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTACAGAAAGGAAAGTATGATGAATA 2037
Qy      661 TYrMeCYsSeRPhELeuPheAsnLeuAsnAspPheValIleAspAlaThRArgLYs 680
Db      2038 TACATGTGACCTTCTGTTCATCTTGAACAATATTTGTGTGAGATGCAACCGGAGAG 2097
Qy      681 GIYAsnLYsLIeArgPheAlaAsnHISeRValAsnProAsnCYSTYrAlaLYsValMeC 700
Db      2098 GGTAACAAAATTCCTTTTGGCAAAATCAATCGGTAAATCCAACTGCTATGCAAAAGTTATG 2157
Qy      701 MecValAsnGIYAspPhISaRGIleGIYLIePhEAlaLYsArgValaIleGInThRGIYGIu 720
Db      2158 ATGTTAAAGGTGATACAGGATAGATATTTTTCGAAGAGGCCATTCAGACTGGGAGAA 2217
Qy      721 GIuLeuPhePheAspTYrArgTYrSeRGIuAlaAspAlaLeuLYsTYrValGIYIleGIu 740
Db      2218 GAGCTGTTTGTATTAAGATTAAGCAAGCTGATCCCTGAAGATATGTGTGGCATGGA 2277

```

Qy 741 ArgGluMetGluIlePro 746  
|||||  
Db 2278 AGAGAAATGGAATCCCT 2295

RESULT 9  
ABX76182  
ID ABX76182 standard; DNA; 2576 BP.  
XX  
AC ABX76182;  
XX  
DT 02-APR-2003 (first entry)  
XX  
DE Lung cancer-associated polynucleotide #54.  
XX  
KW Lung cancer-associated polynucleotide; gene; de; cytosolic; emphysema;  
antiflammatory; antisthmatic; non-small cell lung cancer; atelectasis;  
small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX  
OS Unidentified.  
XX  
PN WO200286443-A2.  
XX  
PD 31-OCT-2002.  
XX  
PF 18-APR-2002; 2002WO-US012476.  
XX  
PR 18-APR-2001; 2001US-0284770P.  
PR 10-MAY-2001; 2001US-0290492P.  
PR 09-NOV-2001; 2001US-0339245P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 29-NOV-2001; 2001US-0343370P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
PA (BOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Aziz N, Murray R;  
XX  
DR WPI: 2003-093161/08.  
DR P-PSDB; ABUS6461.  
XX  
PT Detecting a lung cancer-associated transcript in a cell from a patient  
for treating lung cancer, by contacting a biological sample from the  
PT patient with a polynucleotide that exhibits increased or decreased  
expression in lung cancer.  
XX  
PS Claim 22; Page 231-232; 453pp; English.  
XX  
CC The invention relates to a method for detecting a lung cancer-associated  
transcript in a cell from a patient, comprising contacting a biological  
CC sample from the patient with a polynucleotide that selectively hybridizes  
to a sequence that is at least 80 % identical to a gene that exhibits  
CC increased or decreased expression in lung cancer samples. Lung cancer-  
associated polynucleotides and polypeptides are used for identifying a  
CC compound that modulates a lung cancer-associated polypeptide, for  
CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
CC cancer in a patient and for treating a mammal having lung cancer by  
administering a modulatory compound identified. The methods are useful  
CC for treating lung cancer, such as small cell lung cancer, non-small cell  
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
CC for diagnostic purposes and as targets for screening for therapeutic  
CC compounds that modulate lung cancer, such as antibodies. Sequences  
CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the  
invention  
SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,71e-301 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 2  
Query Match: 99.7% Indels: 0  
DB: 8 Gaps: 0

US-10-773-302-2 (1-746) x ABX76182 (1-2576)

Qy 1 MetGlyGlnThrGlyIleValSerGlyIleProValCysTrpArgValLys 20  
|||||  
Db 58 ATGGCCAGACTGGGAAGAAATCTGAGAGGAGCAAGATTGTTGGCGGAACGGTAAAA 117

Qy 21 SerGlyIleThrGlyLeuArgGlnLeuLysArgPheArgValAlaArgValLysSer 40  
|||||  
Db 118 TCAGAGTCAATGCGACATCGACAGCTCCAGAGGTTCCAGAGCTGATGGAATGAAGT 177

Qy 41 MetPheSerSerAlaValGlnLysIleLeuGluArgThrGlnIleLeuAngGlnLysTrp 60  
|||||  
Db 178 ATGTTTATGTTCCATTCCTCAGAAATTTTGGAAAGACGAAATCTTAAACCAAGATGG 237

Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerIleLeuArgGlyThr 80  
|||||  
Db 238 AAACAGCGAAGAAATACAGCTGTCACATCTGAGCTTCTGAGAGCTCATTGGCGGAGACT 297

Qy 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
|||||  
Db 298 AGGAGTGTTCGGTGCACAGTACCTGGATTTTCCACAGAGCATCCATTAAAGACT 357

Qy 101 LeuAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnAlaAspPhe 120  
|||||  
Db 358 CTGAATCAGTTGCTTCAATGACCATTAATGTTCTGGCTCCCTCCACAGCAAAATTTT 417

Qy 121 MetValGluAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
|||||  
Db 418 ATGTGTGAAGATGAATCACTGTTTATCATTAATCTTATATGGAGATGAAGATTATGAT 477

Qy 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrArgGlyLysValHisGlyAsp 160  
|||||  
Db 478 CAGATGTGATCTTCAATTAAGAACTAATATAAAATTAATATGGAAGAAAGTACAGGGGAT 537

Qy 161 ArgGlyCysGlyPheIleAsnAspGluIlePheValGluLeuValAlaAlaLeuGlyGln 180  
|||||  
Db 538 AGAGAAATGTGGGTTTATTAATGATGAATTTTGTGAGATTGGTAATGCCCTTGGTCAA 597

Qy 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
|||||  
Db 598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657

Qy 201 LysAspLeuGluLysAspHisArgAspAspLysGluSerArgProAlaGlySerPheProSer 220  
|||||  
Db 658 AAAGATCTGGAGGATCCAGAGATGAAGAAAGCGGCCCACTCGAAATTTCTTCT 717

Qy 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluLysLeu 240  
|||||  
Db 718 GATTAATTTTGGAGGCCATTTCTCAATGTTTCCATTAAGGCAACAGAAAGAACTA 777

Qy 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProGluCys 260  
|||||  
Db 778 AAGGAAAAATTAAGAAATCAACCAAGAGCTCCCAAGCGGCACTTCTCCGAAATGT 837

Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
|||||  
Db 838 ACCCCCAACATGATGAGCAAAAGCTTAATCTGTTACAGAGGCAAAAGCTTAACATCC 897

Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
|||||  
Db 898 TTTATAGCGCTTTCTGAGCGATGTTTAAATATGATGCTTCTCTCAATCTTTTCAT 957

Qy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
|||||  
Db 958 GCAACACCAACCACTTAATTAAGCGGAGAAACAGAAACAGCTTACCAACAACTTGT 1017

QY 321 G|YProGlnCySerTyrGlnHisLeuGluGluValAlaLeuPheAlaIleValLeuTyrAla 340  
 DB 1018 GGAACCAAGTGTACAGACATTTGGAGGAGCAAAAGATTGCTGCTCTCTACCGCT 1077  
 QY 341 G|UArgIleLeuThrProProIlybArGProGlyGlyArgArgArgIlyArgLeuProAn 360  
 DB 1078 GAGCGGATTAAGACCCCAACCAAAAGTCAGAGAGCGCGAGAAAGAGACGGCTCCCAAT 1137  
 QY 361 AAmSerSerArgProSerThrProThrIleAenValLeuGluSerIlybArPheThrAspSer 380  
 DB 1138 AACAGTAGAGAGGCCAGACCCCAACATTAATGTGCTGGAATCAAAAGATACAGACACT 1197  
 QY 381 AASPArgIleAlaGlyThrGlyThrGlyGlyGluAsnAsnAspIlybGluGluGlyIlyb 400  
 DB 1198 GATTAGGAGAGCAGGAGCTGAACCGGAGGAGAGACATGATTAAGAGAGAGAGAGAG 1257  
 QY 401 IyAspArgIlybThrSerSerSerSerGluAlaAsnSerArgCySerGlnThrProIleIlybMet 420  
 DB 1258 AAAGATGAACCTTCAGAGCTCCTCTGAGCAAAATTCCTCGGTCAAAACCAATTAAGATG 1317  
 QY 421 IySPProAnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 DB 1318 AAGCCAAATATTAAGACCTCTCAGAAATGAGAGTGGAGTGGCTGAGACCTCAATGTTT 1377  
 QY 441 ArgValIleuIleGlyThrTyrTyrAspAsnPheCyAlaIleAlaArgLeuIleGlyThr 460  
 DB 1378 AAGTTCCTCATGGCACTTACTATGCAATTTCTGTGCAATGTCTGAGCTGAGTAAATGGAGCC 1437  
 QY 461 IySPThrCyAspArgIleValTyrGluPheArgValIyGluSerSerIleIleAlaProAla 480  
 DB 1438 AAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAAGATCTAGCATCTACTCTCAGCT 1497  
 QY 481 ProAlaGluAspValAspThrProProArgIlybIlybIyAspGlybHisIleArgLeuTyrAla 500  
 DB 1498 CCGGCTGAGAGTGGATCTCTCTCCAGGAAAAAGAGAGAAACACCGGTTTGGGGCT 1557  
 QY 501 AlaHisCyArgIlybIleGlnLeuIlybIyAspArgIySerSerAsnHisIleValTyrAsnTyr 520  
 DB 1558 GCACCTGCAAGAAAGATACAGCTGAAAAAGAGCGGCTCTCTTAACCAATGTTTACACTAT 1617  
 QY 521 GlnProCyAspAspHisProArgGlnProCyAspSerSerCySPProCyValIleAlaGln 540  
 DB 1618 CAACCTGTGATCAATCCAGCGAGCTGTGTGACATGTGCTGCTGTGTGTATACACAA 1677  
 QY 541 AsnPheCySerIlybPheCySerGlnCySerSerSerGluCySerGlnAsnArgPheProGlyCyS 560  
 DB 1678 AATTTTGTGAAAAATTTTGTCAATGTAGTTCAAGATGTCAAAACCGCTTCCGGAGATGC 1737  
 QY 561 ArgCybIyValAlaGlnCyAsnThrIlybGlnCySPProCySPTyrLeuAlaValArgGluCyS 580  
 DB 1738 CGCTGCAGAGCAGAGTGCACACCAAGCAAGTCCCGTGTACCTGGCTGTCCGAGAGTGT 1797  
 QY 581 AASPProAspLeuCyLeuThrCySerGlyAlaAlaAspHisSTPASPserIlybAsnValSer 600  
 DB 1798 GACCTGACCTGTGTCTTACTTGTGTGAGCCGCTGACCAATGGGAGAGTAAATATGTGCTC 1857  
 QY 601 CybIyAsnCySerSerIleGlnArgGlySerIlybHisIleLeuLeuAlaProSerAsp 620  
 DB 1858 TGCAAGAACTGACGATTCAGCGGGGCTCCAAAAGCAATCTATCTCGGCACCATCTGAC 1917  
 QY 621 ValAlaGlyTyrPglyIlePheIleIyAspProValGlnIlybAsnGluPheIleSerGlu 640  
 DB 1918 GTGGAGGCTGGGGATTTTATCAAGATCTGTGCAAGAAAAAGAAATTCATTCACAGAA 1977  
 QY 641 TyrCySerGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyIlybValTyrAspIlyS 660  
 DB 1978 TACTGTGAGAGATTAATTTCTCAATGAGTGCAGAAAGGAGAAAGTGTATATATAA 2037  
 QY 661 TyrMetCySerPheLeuPheAsnLeuAsnAsnAspPheValAlaAspAlaThrArgIlyS 680  
 DB 2038 TACATGTGCAGCTTTCTGTTCAACTTGAACAATGATTTTGTGTGAGATGCAACCGCAAG 2097  
 QY 681 GlyAsnIlybIleArgPheAlaAsnHisSerValAsnProAnCyTyrAlaIlybValMet 700

DB 2098 GGTACAAATTCGTTTGCAAATCATTCGGTAAATCCAAATGCTATGCAGAAATTTAG 2157  
 QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaIyArgAlaIleGlnThrGlyGlu 720  
 DB 2158 ATGGTTAACGGTATCACAGATAGGATTTTGTCCAGAGAGCCATCCAGCTGGCGAA 2217  
 QY 721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuIySerTyrValGlyIleGlu 740  
 DB 2218 GAGCTGTTTGTGATTAAGATACAGACGAGCTGATGCCCTGAAATATGTGGCATCGAA 2277  
 QY 741 ArgGluMetGluIlePro 746  
 DB 2278 AGAGAAATGGAATTCCT 2295  
 RESULT 10  
 ADK66988  
 ID ADK66988 standard, DNA; 2576 BP.  
 XX  
 AC ADK66988;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Gene #78 for inhibitory RNA to manipulate stem cell phenotype.  
 XX  
 KW de; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
 XX pluripotent stem cell.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003068961-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-GB0005479.  
 XX  
 PR 13-FEB-2002; 2002GB-00003359.  
 PR 13-FEB-2002; 2002GB-00003387.  
 XX  
 PA (AXOR-) AXORDIA LTD.  
 XX  
 PI Andrews P, Walsh J, Gokhale P;  
 XX  
 WP1; 2003-697528/66.  
 XX  
 PT New inhibitory RNA molecule having double stranded RNA molecules, useful  
 PT for manipulating the phenotype of stem cells, preferably pluripotent  
 PT stem cells.  
 XX  
 PS Disclosure; SEQ ID NO 78; 157bp; English.  
 XX  
 CC The invention relates to an inhibitory RNA (RNAi) molecule derived from a  
 CC nucleic acid molecule comprising a defined nucleic acid sequences given  
 CC in the specific acid or a sequence which hybridizes to the sequences and  
 CC encodes a Notch signaling target gene or which is a degenerate as a  
 CC result of the genetic code of the sequences. The methods and compositions  
 CC of the present invention are useful for manipulating the phenotype of  
 CC stem cells, preferably pluripotent stem cells. This sequence corresponds  
 CC to one of the nucleic acid molecules of the invention.  
 XX  
 SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 7,71e-301 Length: 2576  
 Score: 4038.00 Matches: 744  
 Percent Similarity: 99.7% Conservative: 0  
 Best Local Similarity: 99.7% Mismatches: 2  
 Query Match: 99.7% Indels: 0  
 DB: 10 Gaps: 0  
 US-10-773-302-2 (1-746) x ADK66988 (1-2576)  
 QY 1 MetGlyGlnThrArgIlybIySPserGluIyGluProValCySTPArgGlybArgValIlyS 20

Db 58 ATGGCCAGACTGGAGAAATCTGAGAGGAGACAGTTTGTGGCCGAGCCGTAA 117  
Qy 21 SerGluYrMetArgLeuArgGlnLeuYrPheArgAlaAspGluValIleYrSer 40  
Db 118 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTCACAGAGCTGATGAAAGTAAAGGT 177  
Qy 41 MetPheSerSerAsnArgGlnIleuGlnArgThrGlnIleLeuSerGlnIleYrPhe 60  
Db 178 ATGTTTGGTTCATTCGTCAAGAAATTTGGAAAGAACGAAATCTTAAACAGAAATGG 237  
Qy 61 LysGlnArgArgGlnIleGlnProValHisIleuThrSerValSerSerLeuArgGlyThr 80  
Db 238 AAACAGCGAGAAAGAAATACAGCTGTGACATCTGACCTTCTGTAGAGTCAATTCGGGGAGCT 297  
Qy 81 ArgGluCysSerValIleThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100  
Db 298 AGGAGGTGTGGTGACCGAGTGAAGTGGATTTTCCACACAGATCATCCATTAAAGACT 357  
Qy 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnIleAsnPhe 120  
Db 358 CTGAATGAGTTGCTTCACTAGTACCCATATGTAATTCCTGTCTCCCTCAACAGAAATTTT 417  
Qy 121 MetValGluAspGluThrValIleuHisAsnIleProGlyrMetGlyAspGluValIleuAsp 140  
Db 418 ATGGTGAAGAGTGAACCTGTTTTCATTAACATTCCTTATATGGAGATGAAGTTTATGAT 477  
Qy 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnYrAspGlyLysValHisGlyAsp 160  
Db 478 CAGAGTGTACTTTTCATTGAAAGACTAATTAATAATTTGATGGGAAAGTACAGGGGAT 537  
Qy 161 ArgGluCysGlyPheIleAsnAspGlnIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 538 AGAGATGTGGGTATTAATAGATGAATTTTGTGGAGTGGTGAATGCCCTGTGTCAA 597  
Qy 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnIleArgGlnIleuGln 200  
Db 598 TATATATGATGATGACGATGATGATGACGATCCTGAAAGAAAGAAAGAAAGAG 657  
Qy 201 LysAspLeuGlnAspHisArgAspAspLysGlnIleAspProProArgLysPheProSer 220  
Db 658 AAAGATCTGAGAGATCACCGAGATATTAAGAAACCGCCACCTCCGAAATTTCTTCT 717  
Qy 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnIleu 240  
Db 718 GATAAAATTTTGGAGGCCATTTCTCCATGTTTCCAGATTAAGGCGACAGCAAGAACTA 777  
Qy 241 LysGlnIleYrLysGlnIleuThrGlnIleuProGlyAlaLeuProProGluCys 260  
Db 778 AAGGAAATATTAAGAACTCACCGAACAGCGCTCCAGCGGCACTTCTCTGTAATGT 837  
Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnIleSerLeuHisSer 280  
Db 838 ACCCCCAATATGATGAGACCAATATCTGTTCAAGAGAGCAAAAGCTTACATCC 897  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLysYrAspCysPheLeuHisIleProPheHis 300  
Db 898 TTTCAATAGCTTTCTGTAGGCGATGTTTAATATGACTGCTTCTCATATCTTTTCAAT 957  
Qy 301 AlaThrProAsnThrTyrIleValArgLysAsnThrGlnIleValAlaLeuAspAsnLysProCys 320  
Db 958 GCACACCCCACTTAATTAAGCGAAGAACACAGAAACAGCTCTAGACAAACAACTTGT 1017  
Qy 321 GlyProGlnCysTyrGlnHisIleuGlnIleValLysGluPheAlaAlaIleuThrAla 340  
Db 1018 GGACCAAGGTGTTACAGCATTTGAGGGAGCAAAAGAGTTTGTCTGCTCTCAACGGCT 1077  
Qy 341 GlnArgIleLysThrProProLysArgProGlyArgArgArgArgIleArgLeuProAsn 360  
Db 1078 GAGCGGATTAAGAACCCCAACAAACGTCCAGAGAGCGCGCAAGAGAGAGCGGCTCCCAT 1137  
Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysPheThrAspSer 380

Db 1138 AACATGACAGCGCCACGACCCCAACATTATGTCTGTAATCAAGATACAGACGT 1197  
Qy 381 AspArgGlnAlaGlyThrGlnIleuGlyGlnIleAsnAsnAspLysGlnIleuGlnIleuLys 400  
Db 1198 GATAGGAGACAGAGACTGAACCGGGGGAGAGAAACATGATTAAGAAAGAAAGAGAG 1257  
Qy 401 LysAspGlnIleThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnIleThrProIleYrMet 420  
Db 1258 AAAGATGAACCTTCGAGCTCTCTGAAGCAAAATTCGTGGTGTCAACCAATTAAGATG 1317  
Qy 421 LysProAsnIleGluProProGlnIleuValGlnIleuTrpSerGlyAlaGlnAlaSerMetPhe 440  
Db 1318 AAGCCAAATTTTGAACCTCTGAGATGTGAGATGTGAGTGTCTGAAGCTCAATGTTT 1377  
Qy 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1378 AGAGTCTCATTTGGCACTTACTATGACAAATTCGTGTCCTAGTTCAATTTGGGAGCC 1437  
Qy 461 LysThrCysArgGlnValIleArgIleuPheArgValIleGlnSerSerIleIleAlaProAla 480  
Db 1438 AAACATGTAGACAGGTGATGAGTTTGAAGTCAAGAAATCTAGCATATAGCTCCAGCT 1497  
Qy 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
Db 1498 CCGCTGAGATGTGATCTCTCCAGAGAAAGAAAGAAAGAAACACCGGTTGTGGGCT 1557  
Qy 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValIleYrAsnTyr 520  
Db 1558 GCACACTGCAAGAAAGATACAGCTGAAGAAAGAGAGCGCTCCCTCAACCAATGTTTCAACTAT 1617  
Qy 521 GlnProCysAspHisPheProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1618 CAACCTGTGATCATCCAGCGAGCTTGAGACGTTGAGCGCTTGTGTGTGATAGCAAA 1677  
Qy 541 AsnPheCysGlnLysPheCysGlnCysSerSerSerGlnCysGlnAsnArgPheProGlyCys 560  
Db 1678 AATTTTGTGAAAGTTTGTGTCAATGTATGTCAGAGTCAAAACCGCTTTCCGGAGATGC 1737  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrIleuAlaValArgGlnCys 580  
Db 1738 CGCTGCAAGACAGTGCACACACCAAGACAGCGCCGTCACTGAGCTGTCGAGAGATGT 1797  
Qy 581 AspProAspLeuCysLysLeuThrCysGlyValAlaAspHisIleThrAspSerLysAsnValSer 600  
Db 1798 GACCTTCACTCTGTCTTACTTGTGAGCGCGTGAACCATTTGGACAGATTAATAATGTGTC 1857  
Qy 601 CysLysAsnCysSerIleGlnArgGlySerIleLysHisIleuLeuAlaProSerAsp 620  
Db 1858 TGCAGAACTGCAATATTCAGCGGGGCTCCAAAAGCATATTTGCTGGCACATCGAC 1917  
Qy 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGln 640  
Db 1918 GTGGCAGGCTGGGGATTTTATCAAAAGATCTGTGACAGAAATGAATTCATCTCAGAA 1977  
Qy 641 TyrCysGlnIleGlnIleSerGlnAspGlnAlaAspArgArgGlyLysValIleYrAspLys 660  
Db 1978 TACGTGTGAGATTAATTTCTCAAGATGAAGTCAAGAAAGGAGAAAGTTATGATTA 2037  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgLys 680  
Db 2038 TACATGTGCACTTCTGTTCACTTGAACATGATTTTGTGTGAGATGCAACCGGAG 2097  
Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
Db 2098 GGTAACAAAATTCGTTTGCAAATCATTCGGTAATCCAATCTGTCGCAAAAGTTATG 2157  
Qy 701 MetValAsnGlnYrAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
Db 2158 ATGGTTAACTGTATCAACAGATGATGATTTTTCCAAGAGAGCATTCAGACTGGCGGAA 2217  
Qy 721 GlnLeuPhePheAspYrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGln 740  
Db 2218 GAGCTGTTTGTGATTAACATACAGCAAGCTATGCCCTGGAAGTATGTGGCATGAA 2277



QY 741 ArgGluMetGluIlePro 746  
|||||  
Db 2278 AGAGAAATCGAAATCCT 2235

RESULT 11  
ID ADN39391  
ADN39391 standard; cDNA; 2576 BP.

AC ADN39391;  
XX  
XX 17-JUN-2004 (first entry)  
DT  
XX  
DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:875.  
XX  
XX Human; differential expression; cancer; angiogenic disorder;  
KM fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KM inflammatory disease; autoimmune disease;  
KM retinal neovascularization syndrome; scarring; uterine fibroid;  
KM detection; diagnosis; prognosis; drug screening; drug targeting;  
KM wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
KM vulnerrary; gene therapy; vaccine; gene; ss.

OS Homo sapiens.  
XX  
XX MO2003042661-A2.  
PD 22-MAY-2003.

XX 13-NOV-2002; 2002WO-US036810.  
PF  
XX 13-NOV-2001; 2001US-0350666P.  
PR 21-NOV-2001; 2001US-0332464P.  
PR 29-NOV-2001; 2001US-0343933P.  
PR 03-DEC-2001; 2001US-0335394P.  
PR 14-DEC-2001; 2001US-0340376P.  
PR 08-JAN-2002; 2002US-0347211P.  
PR 10-JAN-2002; 2002US-0347349P.  
PR 08-FEB-2002; 2002US-0355250P.  
PR 13-FEB-2002; 2002US-0356714P.  
PR 20-FEB-2002; 2002US-0359077P.  
PR 29-MAR-2002; 2002US-0368009P.  
PR 04-APR-2002; 2002US-0370110P.  
PR 12-APR-2002; 2002US-0372246P.  
PR 05-JUN-2002; 2002US-0386614P.  
PR 16-JUL-2002; 2002US-0396839P.  
PR 22-JUL-2002; 2002US-0397757P.  
PR 09-SEP-2002; 2002US-0409450P.

XX (BOSB-) BOS BIOTECHNOLOGY INC.  
PA  
XX Afar D, Aziz N, Gineburg WM, Gish KC, Glynn R, Hevez PA;  
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnick A;  
XX WPI. 2003-468649/44.  
DR P-ESDB; ADN39392.

XX Determining the presence or absence of a pathological cell in a patient,  
PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
PT a nucleic acid in a biological sample.

XX Claim 8; SEQ ID NO 875; 1385bp; English.  
PS  
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;

CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularization syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a nucleic acid sequence of the invention.  
XX

SEQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 7,71e-301 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 2  
Query Match: 99.7% Indels: 0  
DB: Gaps: 0

US-10-773-302-2 (1-746) x ADN39391 (1-2576)

QY 1 MetGlyInThrGlyValSerGluValProValCysTrpArgValValys 20  
Db 58 ATGGCCAGACTGGGAGAAATCTGAGAGACCGAGTTGTGGCGAAGCGTAA 117

QY 21 SerGluTyrMetArgLeuArgGluLeuValArgPheArgAlaAspGluValLysSer 40  
Db 118 TCGAGTACATGCGATGACGATGACGATGACGATGACGATGACGATGACGAT 177

QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
Db 178 ATGTTTACTTCATTCGTCAGAAAATTTGGAAAGAACGAAATTTAAACCAAGATG 237

QY 61 LysGlnArgGlnGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 238 AAACAGCAGAAAGATACAGCTGTCGACCTGCTGCTGCTGCTGCTGCTGCTGCT 297

QY 81 ArgGluCysSerValIleThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 298 AGGAGTGTTCGTCGACGATGACGATGATGATGATGATGATGATGATGATGATGAT 357

QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnAsnPhe 120  
Db 358 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417

QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
Db 418 ATGGTGAAGATGAACGTTTACATTAATCTTCTATATGGGAGATGAAGTTTAAAT 477

QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyValValHisGlyAsp 160  
Db 478 CAGATGTGATCTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 537

QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAlaLeuGluLysGln 180  
Db 538 AGAGATGTGGCTTATTAATGAATGAATTTTGTGAGTTGTGAAATGCCCTTGCTAA 597

QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
Db 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657

QY 201 LysAspLeuGluAspPheIleArgAspAspLysGluSerArgProProArgLysPheProSer 220  
Db 658 AAAGATCTGGAGATCACCGAGATGAATGAAGAGCCGACCTGGAAATTTCTCTCT 717

QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
Db 718 GATTAATTTTGGAGGCGATTTCTCAATGTTTCCAGTAAAGGACACGAGAAAGAACTA 777

QY 241 LysGluLysTyrLysGluLeuThrGluGlnGluLeuProGlyAlaLeuProProGluCys 260  
Db 778 AAGGAAAAATTAAGAACTACCGAAGAGCTCCAGGCGGACCTTCCTGAAATGT 837



Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	7, 71e-301	Length:	2576
Score:	4038.00	Matches:	744
Percent Similarity:	99.7%	Conservative:	0
Best Local Similarity:	99.7%	Mismatches:	2
Query Match:	13	Indels:	0
DB:	13	Gaps:	0

US-10-773-302-2 (1-746) x ADR25182 (1-2576)

```
QY 1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20
DB 58 ATGGGCGCAGACTGGGAGAAATCTGAGAAAGGACCACTTGTGGCGGAAAGCGTGAATA 117
QY 21 SerGlyLysMetArgLysLysGlnLysLysArgPheArgArgAlaAspGluValLysSer 40
DB 118 TCAGAGTACATCGACTGAGACAGGCTCAAGAGGTTCAAGCAGCTGATGAACTAAAGAT 177
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluValGlnIleLeuAsnGlnGluTrp 60
DB 178 ATGTTTAGTTCCAATCGTCAGAAATTTTGAAAGAACCGAATCTTAAACCAAGATCG 237
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerLeuArgGlyThr 80
DB 238 AAACGCGAAGATACAGCTGTGACATCTGACTCTGTGAGCTCATTCGCGGAGCT 297
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100
DB 298 AGGAGTGTTCGCTGACCAAGTACCTTGATTTCCAAACCAAGTATCCCATTAAGACT 357
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnAsnPhe 120
DB 358 CTGATGCGAGTTCAGTACGATCCCAATATGATTCCTGCTCCCTACAGCAGAAATTT 417
QY 121 MetValGluAspGluThrValLeuHisIleProLysMetGlyAspGluValLeuAsp 140
DB 418 ATGGGGAAGATGAACGTTTACATTAACATTCCTATATGAGATGAATGATTTAGAT 477
QY 141 GlnAspGlyThrPheIleGluGlnLeuIleLysAsnLysArgGlyLysValHisIleLys 160
DB 478 CAGATGTGACTCTTCATTCAGAACTAAATTAATTAATGATGCGAAAGTACACGGGAT 537
QY 161 ArgGlyCysGlyPheIleAsnAspGlnIlePheValGluLeuValAsnAlaLeuGln 180
DB 538 AGAGATGTGGGTTATATAATGATGAATTTTGTGAGGTGTGTGAATCCCTGTGCTCA 597
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnLysGln 200
DB 598 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
QY 201 LysAspLeuGluAspHisIleArgAspAspLysGlnSerArgProProArgLysPheProSer 220
DB 658 AAAGATCTCGAGGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTTCT 717
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGlnLeu 240
DB 718 GATAAAATTTTGGAGGCACTTCTCAATGTTTCCAGATTAAGGCAAGCAGCAGCAAGACTA 777
QY 241 LysGlnLysLysLysGlnLeuThrGlnGlnGlnLeuProGlyValAlaLeuProProGlu 260
DB 778 AAGGAAATATATAAGAACTACCAAGAACGAGCTCCAGGCGCACTTCTCTTAATAT 837
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnIleSerLeuHisSer 280
DB 838 ACCCCCAATATGATGAGCAACCAATGCTAAATCTGTTCAAGAGAGCAAGCTTACACTC 897
QY 281 PheHisIleThrLeuPheCysArgArgCysPheLysLysLysPheLeuHisIleProPheHis 300
DB 898 TTTCATAGCGCTTTCTGATGCGGATGTTTAAATATGATGCTCTCATCATCTTTTCAT 957
QY 301 AlaThrProAsnThrLysLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320
```

```
DB 958 GCAACACCCCAACCTTATTAAGGGAAGAACACAGAAACAGCTCTTGAACAACAACCTTCT 1017
QY 321 GlyProGlnCysTrpGlnHisIleLeuGlnValAlaLysGlnPheAlaAlaLeuThrAla 340
DB 1018 GGACCAAGTGTTCACCACTTGTGAGGAGCAAGGAGTTTGTGCTGTCTCTCAACCGCT 1077
QY 341 GlnArgIleLysThrProProLysArgProGlnGlyValArgArgArgGlyArgLeuProAsn 360
DB 1078 GAGCGAATTAAGACCCCAACAAACGTCAGAGGCGCCAGAAAGGACGCGCTTCCAT 1137
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnLysSerLysAspThrAspSer 380
DB 1138 AAGATGACAGGCCCCAGACACCCCACTTAATGTGCTGGAATCAAGATACACACT 1197
QY 381 AspArgGlnAlaGlyThrGluThrGlyGlyGluAsnAspLysGlnGlnLys 400
DB 1198 GATAGGAAGCAGGAGTGAACCGGGGAGAGAACTATGATTAAGAAAGAAAGAAAG 1257
QY 401 LysAspGlnThrThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420
DB 1258 AAAGATGAACCTTCAGCTCTCTGAGCCAAATTCCTGCTGTCACCAATTAAGATG 1317
QY 421 LysProAsnIleGlnProProGlnAsnValGluTrpSerGlyAlaGlnAlaSerMetPhe 440
DB 1318 AAGCAATATTTGAACCTCTGAGAAATGTGAGTGAATGATGCTGAAGCTCATGCTTT 1377
QY 441 ArgValLeuIleGlyThrLysTrpAsnAspPheCysAlaIleAlaArgLeuIleGlyThr 460
DB 1378 AAGTCTCATTTGGCACTTACTATGACAAATTTCTGTGCAATTCGATGAGTTAATGGAGC 1437
QY 461 LysThrCysArgGlnValLysGlnPheArgValLysGlnLysSerIleIleAlaProAla 480
DB 1438 AAAACATGTAGCAGGTGTATGATTAAGATCAAGAAATCATGATCATATGCTCAGCT 1497
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisIleArgLeuTrpAla 500
DB 1498 CCGCTGAGATGTGATGATCTCTCCAGAGAAAGAAAGAGAAACACCGGTGTGGGCT 1557
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValLysThrArg 520
DB 1558 GCACACTGCAAGAAAGATACAGTGAATAAGACCGCTCTCTTAACATGTTTACAACATAT 1617
QY 521 GlnProCysAspAspPheProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB 1618 CAACCTGTGATCATCCACGAGCGCTGTGACAGTTGTGTGCTGTGTGATGATACACA 1677
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560
DB 1678 AATTTTGTGAAGATTTTGTCAATGTATGTTCAGATGTCAAAACCGCTTCCGGGATGC 1737
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysLysLysValAlaArgLysCys 580
DB 1738 CGCTGCAAAAGCAGGTGCAACCAAGCAGAGGCCGTCTACCTGCTGTCCGAGAGTGT 1797
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAlaAspHisIleTrpAspSerLysAsnValSer 600
DB 1798 GACCTGACCTCTGTCTTAATCTGTGAGCCCGCTGACATTTGGAGCAGTAAGAAATGTGCC 1857
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleLeuLeuAlaProSerAsp 620
DB 1858 TGCAGAACTGAGATATTCAGGGGGCTCCAAAGACATCTATCTGTGACACACTTCAAC 1917
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGln 640
DB 1918 GTGGCAGGCTGGGGGATTTTATCAAGATCTGTGCGAAGAAATGAATTCATCTCAGAA 1977
QY 641 TyrCysGlyGlnIleIleSerGlnAspGlnAlaAspAspArgGlyLysValLysTrpAspLys 660
DB 1978 TACTGTGAGAGATTAATTTCTCAAGTGAACCTGACAAAGAGGAAAGTGTATGATAAA 2037
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680
```

Db 2038 TACATGTCAGCTTCTGTTCACTTGAACAATGATTTTGGTGGATGCAACCCGCAAG 2097  
Qy 681 G1yAenlyS1leArphea1Aasn1sSerVal1asnProasnCyS1yZal1ySVal1Met 700  
Db 2098 GGTAAACAAAATTCCTGTTGGCAATCATTCGGTAATCCAAATCTGTATGCAAAAGTTATG 2157  
Qy 701 MetValang1yAerPH1sArx1leg1y1lePhea1alyArGal1a1leg1nTh-G1yGlu 720  
Db 2158 ATGGTTAAACGGTGAATCAACAGATAGATTTTGGCAGAGAGCCATCCAACTGGCCGAA 2217  
Qy 721 Glu1euPhePheArp1yArx1ySerg1n1a1aerPal1eul1yS1yVal1G1y1leg1u 740  
Db 2218 GAGCTGTTGTTGATACAGATACAGCCAGCGTGAATGCCCTGAAGTATGCGCATCGAA 2277  
Qy 741 Arg1u1e1cG1u1lePro 746  
Db 2278 AGAGAAATGAAATCCCT 2295  
RESULT 13  
ADY61813  
ID ADY61813 standard; DNA; 2576 BP.  
XX AC ADY61813;  
DT 19-MAY-2005 (first entry)  
XX DE Human gene marker indicating poor prognosis for breast cancer Seq 129.  
XX KM gene; ds; prognosis; breast cancer; diagnosis; RNA interference;  
KM screening; gene therapy; cytostatic.  
XX OS Homo sapiens.  
XX PN US2005054826-A1.  
XX PD 10-MAR-2005.  
XX PF 18-MAY-2004; 2004US-00848755.  
XX PR 19-MAY-2003; 2003US-0471842P.  
XX PA (ROSE-) ROSETTA INPHARMATICS LLC.  
XX P1 Mao M;  
XX DR WPI; 2005-213102/22.  
XX PT New purified diaphanous-3 (DIAPH3) peptides and nucleic acids, useful as  
PT targets for screening assays designed to identify inhibitors or  
PT modulators of DIAPH3 activity, which are useful for treating breast  
PT cancer.  
XX PS Disclosure; SEQ ID NO 129; 54bp; English.  
XX CC This invention relates to a novel human diaphanous-3 (DIAPH3) protein and  
CC the encoding polynucleotide thereof. Specifically, it refers to the use  
CC of DIAPH3 cDNA in the prognosis of breast cancer, such that the DIAPH3  
CC protein can be used as a target for anti-cancer drugs or for the  
CC identification of molecules with anti-cancer activity. The present  
CC invention provides antibodies that specifically bind to the DIAPH3  
CC protein, as well as methods to identify an agent that modulates the  
CC binding of DIAPH3 to a binding partner. In particular it describes  
CC measuring the expression level of DIAPH3 using hybridization oligos and  
CC comparing this with a control expression level, where a raised expression  
CC level indicates a poor prognosis with respect to breast cancer. As such,  
CC it also provides small interfering RNA oligos that can hybridize to the  
CC DIAPH3 gene and inhibit expression thereof. The peptide and nucleic acid  
CC sequences of this invention are useful as markers to differentiate  
CC patients with a good prognosis from those with poor prognosis for breast  
CC cancer and are useful as targets for screening assays designed to  
CC identify inhibitors or modulators of DIAPH3 activity, which in turn are  
CC used for treating breast cancer. Accordingly, they can be used in gene  
CC therapy and exhibit cytostatic activity. This polynucleotide sequence is

CC a human gene marker that when expression is upregulated in a patient  
CC sample it is correlated with a poor prognosis for breast cancer, given in  
CC an exemplification of the invention.  
XX SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 7,71e-301 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 2  
Query Match: 99.7% Indels: 0  
DB: 14 Gaps: 0  
US-10-773-302-2 (1-746) x ADY61813 (1-2576)  
Qy 1 MetG1yG1nThrG1y1yS1ySerg1u1yG1yProVal1CyS1TPArG1ySArGVal1yS 20  
Db 58 ATGGCCAGACTGGGAGAAATCTGAGAGAGGACCACTTGTGGCGGAGCGTGAATAA 117  
Qy 21 Serg1u1yMetArG1eulArG1n1eul1yArG1rPhea1rGal1aerG1uVal1ySer 40  
Db 118 TCAGAGTACATGCGACTGAGACAGACTCAAGAGTTTCAGACGACTGATGAAGTAAAGAT 177  
Qy 41 MetPheSerSerAanaArG1n1yS1le1eulG1uArGrThrG1u1le1eulAng1nG1uTrp 60  
Db 178 ATGTTTGTTCCAAATCTCGAATAATTTGGAAAGAAACGAAATCTTAACCAAGAAATGG 237  
Qy 61 LyS1nArGrArG1leg1nProVal1H1e1eulThrSerVal1SerSer1eulArG1yThr 80  
Db 238 AAACGCGAAGAGATACAGCCGCTGTCACATCTGAGCTCATTTGCGGGGACT 297  
Qy 81 ArG1uCySerVal1ThrSerArP1eulArPheProThrG1nVal1lePro1eul1yThr 100  
Db 298 AGGAGTGTCCGGTGAACAGTGAATTTTCCACACAAAGTCATCCATTAAGACT 357  
Qy 101 LeuAen1aVal1aSerVal1Pro1leMer1ySer1rPhePro1eulG1n1aSnPhe 120  
Db 358 CTGAATCAAGTGTTCAGTACCAATATATCTTGCTCCCTACAGCAAAATTTT 417  
Qy 121 MetVal1G1u1eP1G1uThrVal1leuH1eAasn1lePro1yThrMetG1yArP1G1uVal1eulArP 140  
Db 418 ATGCTGGAAGATGAATCTGTTTACATACATCTCTATATGGAGATGAAGTTTGAT 477  
Qy 141 GluAerG1yThrPhe1leg1u1eul1e1yAen1yArP1G1yVal1H1eG1yArP 160  
Db 478 CAGATGTAATCTTCAATTAAGAACTAATTAATAAATATATGATGGAAAGTACAGGGGAT 537  
Qy 161 Arg1u1yCserG1yPhe1leAenArP1u1lePheVal1G1u1eulVal1aenAla1eulG1n 180  
Db 538 AGAGAAATGGGTTTAAATGATGAATAATTTTGTGAGTTGTGAATGCCCTTGCTCA 597  
Qy 181 TyzAenArP1eArP1eArP1eArP1eArP1eArP1eArP1eArP1eArP1eArP1eArP1e 200  
Db 598 TATAATGATGATGACATGANTGANTGANTGANTGANTGANTGANTGANTGANTGANTGANT 657  
Qy 201 LySArP1eulG1uArP1e1sArGArP1yS1uSerArP1yProArG1yPheProSer 220  
Db 658 AAAGATCTGAGGATCAACCGAGATGAAGAAACCGCCCACTCGGAATTTCTCTTCT 717  
Qy 221 Asp1yS1lePheG1u1a1le1SerSerMetPheProArP1yS1yThrAla1G1u1eul 240  
Db 718 GATAAATTTTGGAGGCCATTTCTCAATGTTTCAATTAAGGACACAGCAAGAAACTA 777  
Qy 241 LySg1u1yS1y1y1ySg1u1eulThrG1u1nG1n1eulProG1yAla1eulProProG1u1yS 260  
Db 778 AAGAAATATTAAGAATCAACGAAACAGAGCTCCAGCGGCACTTCTCTGAAATGT 837  
Qy 261 ThrProAen1leArP1yProAen1a1ySerVal1G1nArG1u1nG1u1nSer1eulH1eSer 280  
Db 838 ACCCCAAACATAGATGACCAATATCTGTTCAAGAGAGCAAAAGCTTACACTCC 897  
Qy 281 Phe1sThr1eulPheCySArGArG1yPhe1yS1yThrArP1yPhe1eulH1eProPheH1e 300

Db 898 TTTCTATACGCTTTCTGTAAGCGATGTTTAAATATGATGCTTCTTATACCTTTTCAAT 957  
 Qy 301 AAtthrProaenThrTylysaarglysaenThrglyThralaleuapnaenlyProCys 320  
 Db 958 GCACACCCCAACCTTATAGCGGAAGAACACAGAAACAGCTCTAGACCAACCAACTTGT 1017  
 Qy 321 GLYProGlnCysEThrGlnHisleuGlnGlyAlaIysGlnPheAlaAlaAlaIleuThrala 340  
 Db 1018 GAAACACAGTGTATACAGCATTTGGAGGAGCAAGAGATGTGCTGCTCTCTACCGCT 1077  
 Qy 341 GIuArgIleLysThrProProLysArgProGIyGIyArgArgArgGIyArgLeuProaen 360  
 Db 1078 GAGCGGATTAAGACCCCAACCAAAACGTCCAGAGAGCGCGCAAGAGAGAGCGGCTCCCAAT 1137  
 Qy 361 AAnsSerArgProSerThrProThrIleAenValleuGlnSerIlysaThrAspSer 380  
 Db 1138 AACAGTACAGAGCGCCACGCCACCATTAATGTGCTGGAATCAAAAGATACAGACAGT 1197  
 Qy 381 AspArgGluAlaGlyThrGlnThrGlyGlyIysAenAenAspLysGlnGlnGlyLys 400  
 Db 1198 GATAGGGAAGACAGGAGCTGAACCGGGGAGAGAACATGATTAAGAAAGAGAGAGAG 1257  
 Qy 401 LysAspGluThrSerSerSerSerGluAlaAenSerArgCysGlnThrProIleLysMet 420  
 Db 1258 AAAGATGAACCTTCAGACTCTCTGAGAGCAATTCGSGTGTCAAAACCAATTAAGATG 1317  
 Qy 421 LysProaenIleGlnProProGlnAenValGluThrSerGlyAlaGlnAlaSerMetPhe 440  
 Db 1318 AAGCCAAATATTGAACCTCTCTGAGAAATGAGAGTGAAGTGGTGGTGAACCTCAATGTTT 1377  
 Qy 441 ArgValleuIleGlyThrTyThrAspAenPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1378 AAGAGCTCCATGAGCACTTACATATGACATTTCTGTCGCAATGTCTAGAGTTATGGGACC 1437  
 Qy 461 LysThrCysArgGlnValTyArgLysPheArgValLysGlnSerSerIleIleAlaProAla 480  
 Db 1438 AAAAACTGAGACAGGTGTATAGATTAGATCAAAAGATCTAGACATCTACTCAGCT 1497  
 Qy 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrala 500  
 Db 1498 CCGGCTGAGAGATGTGATCTCTCTCAAGAAAGAAAGAGAAACACCGGTTGTGGGCT 1557  
 Qy 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAenHisAlaTyThrAenTy 520  
 Db 1558 GCACACTGCAGAAAGATACAGCTGAAGAAAGAGCGGCTCTCTTACCATGTATTACACATAT 1617  
 Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 1618 CAACCTGTATCATCTCAGCGAGCGCTGTGACAGTTCTGTGCTGTGTATGATGACAA 1677  
 Qy 541 AenPheCysGluLysPheCysGlnCysSerSerSerGlyCysGlnAenArgPheProGlyCys 560  
 Db 1678 AATTTTGTGAAAAGTTTGTCAATGTATGTTCAAGGTGTCAAAACCGCTTCCGGGATGC 1737  
 Qy 561 ArgCysLysAlaGlnCysAenThrLysGlnCysProCysTyThrLeuAlaValArgLysCys 580  
 Db 1738 CGCTGCAAGCAAGGAGCAACACAGAGTCCCGTGTCTACCTGCTGTCCGAGAGTGT 1797  
 Qy 581 AspProaenLysCysLeuThrCysArgIlyAlaAlaAenHisThrAspSerIlysaAenValSer 600  
 Db 1798 GACCTGTACCTGTCTTACTGTGTGAGCGCTGTACCATTTGGAGACATTAATAATGTGCTC 1857  
 Qy 601 CysLysAenCysSerIleGlnArgLysSerIlyLysHisleuLeuLeuAlaProSerAsp 620  
 Db 1858 TGCAGAACTGCAGATTCAGCGGGGCTCCAAAAGCAATCTATTGTGTCACCATCTGAC 1917  
 Qy 621 ValAlaGlyTyThrGlyTyrIlePheIleLysAspProValGlnLysAenGlnPheIleSerGlu 640  
 Db 1918 GTGGCAGAGCTGGGGGATTTTATCAAAAGATCTGTGTCAGAAAATGAATTCATCTCAGAA 1977  
 Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyThrAspLys 660

Db 1978 TACTGTGAGAGATATTCTCAAGATGAGCTGACAGAAAGAGGAAAGTATGATATAA 2037  
 Qy 661 TyrMetCysSerPheLeuPheAenLeuAenAenAspPheValValAspAlaThrArgLys 680  
 Db 2038 TACATGTCAAGCTTCTCTGTCAACTGAAACATGATTTTGTGTGAATGCAACCCGCAAG 2097  
 Qy 681 GlyAenLysIleArgPheAlaAenHisSerValAenProaenCysEThrAlaLysValMet 700  
 Db 2098 GGTAAACAAATTCGTTTTCGCAAAATTCGTAATCCAACTGCTATGCAAAAGTTATG 2157  
 Qy 701 MetValaenGlyAspHisArgIleGlyTyrIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
 Db 2158 ATGTTTAAGGTGATCACAGGATAGTATTTTGGCAAGAGCCATCCAGACTGGCGAA 2217  
 Qy 721 GluLeuPhePheAspTyThrArgTySerGlnAlaAspAlaLeuLysTyValGlyTyrIleGlu 740  
 Db 2218 GAGCTGTTGTGATTAACATACATACAGCGAGCTGATGCCGTAAGTATGTCCGATCGAA 2277  
 Qy 741 ArgGluMetGluIlePro 746  
 Db 2278 AGAGAAATGAAATCTCT 2295  
 Db 2278 AGAGAAATGAAATCTCT 2295  
 RESULT 14  
 ADP95921  
 ID ADP95921 standard; DNA; 2253 BP.  
 AC ADP95921;  
 AC ADP95921;  
 DT 07-OCT-2004 (first entry)  
 DE Human EZH2 (enhancer of zeste homologue 2)-encoding DNA, SEQ ID NO:2.  
 XX Human; EZH2; enhancer of zeste homologue 2; chromosome 7q35;  
 KW histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
 KW inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
 KW lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
 KW prophylaxis; prevention; diagnosis; cytostatic; gene; ds.  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT 1..2253  
 CDS /tag=a  
 FT /partial  
 FT /product= "EZH2 (enhancer of zeste homologue 2)"  
 FT /note= "No stop codon given"  
 XX  
 PN WO2004058969-A1.  
 XX  
 PD 15-JUL-2004.  
 XX  
 XX 22-DEC-2003; 2003WO-JP016417.  
 PF  
 XX 24-DEC-2002; 2002JP-00373144.  
 PR  
 XX (TAKES ) TAKEDA CHEM IND LTD.  
 PA  
 XX HIKICHI Y, NISHIZAWA S;  
 PI  
 XX  
 XX  
 DR WPI; 2004-525885/50.  
 DR P-PsDB; ADP95920.  
 XX  
 XX Novel prophylactic and therapeutic agent of cancer, having compound or  
 PT its salt that inhibits activity of enhancer of zeste Drosophila homologue2  
 PT protein, useful for treating cancer such as lung cancer, breast cancer.  
 XX  
 PS Disclosure: SEQ ID NO 2; 102pp; Japanese.  
 CC The invention relates to an agent for the prophylaxis and therapy of  
 CC cancer comprising a compound which inhibits the activity or expression of  
 CC human EZH2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
 CC EZH2 inhibitors include anti-EZH2 antibodies and EZH2 antisense nucleic  
 CC acids. The invention also relates to a kit for screening for inhibitors

CC of EZH2 activity or expression. EZH2 is a component of a complex with  
CC histone methyltransferase activity which methylates histone H3 on lysine  
CC residues at position 9 and/or 27. It has been found to be deregulated in  
CC various tumours, and acts as a mediator of tumorigenesis, being able to  
CC reverse repression of cyclin A which in turn mediates cell cycle  
CC advancement. The agent of the invention can be used for preventing,  
CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
CC in drug screening. The agent of the invention is safe and less toxic than  
CC prior art cancer prophylactic/therapeutic agents. The present sequence  
CC represents DNA encoding human EZH2.

XX Sequence 2253 BP; 717 A; 449 C; 546 G; 541 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	2,45e-300	Length:	2253
Score:	4030.50	Matches:	745
Percent Similarity:	99.28	Conservative:	0
Best Local Similarity:	99.28	Mismatches:	1
Query Match:	99.54	Indels:	5
DB:	12	Gaps:	1

US-10-773-302-2 (1-746) x ADP95921 (1-2253)

QY 1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
DB 1 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTTGTCGGAGAACGTGTAAAA 60  
QY 21 SerGlyuTyrMetArgLysLeuArgGlnLeuLysArgPheArgAlaAspGlyValLysSer 40  
DB 61 TCAAGATACATGCACTGACAGACCTCAAGAGTTCAACAGAGCTGATGAAGTAAAGGT 120  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
DB 121 ATGTTTATGTTCCAAATCGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATGG 180  
QY 61 LysGlnArgArgGlnLeuProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 181 AAACGCGAAGATGACAGCTGTGTGACATCTGACCTTGATGAGCTCATTCGCGGAGACT 240  
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 241 AGGAGGTGTGCTGACCAAGTACTGACTGATTTTCCAAACAGATCATCTCCATTAAAGCT 300  
QY 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnAlaAspHe 120  
DB 301 CTGATGAGAGTTGCTTCAAGTACCATTAATGTATTTCTGTCTCCCTCAGAGAAATTTT 360  
QY 121 MetValGluAspGlyThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 361 ATGGGGAAGATGAACCTGTTTACATTAACATTCCTTATATGGAGATGAAGTTTAAAT 420  
QY 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnTyrAspGlyLysValHisIleGlyAsp 160  
DB 421 CAGAGTGTGATCTTATGAGAACTTAATAAAATTTATGATGGAAAGTACACGGAGAT 480  
QY 161 ArgGlyCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 481 AGAGAAATGTGGGTTATTAATGATGAATTTTGTGGAGTTGTGATGCCCTTGCTCA 540  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGlnLysGln 200  
DB 541 TATATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 601 AAAGATCTGGAGAGTACACCGAGATATTAAGAAACCGCCCACTCGAAATTTCTCTCT 660  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGlnLeu 240  
DB 661 GATAAATTTTGAAGCCATTTCTCATATGTTTCCAGATTAAGGGCAGCAGAAAGACTTA 720  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnLeuProGlyAlaLeuProProGluCys 260

DB 721 AAGGAAAAATTAAGAACTCACCGAACAGAGCTCCCAAGCAGCTTCTCTCGAATGT 780  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 781 ACCCCCAACATATGATGAGCAACCAATGCTTAATCTGTTCAGAGAGCAAGCTTACCTCC 840  
QY 281 PheHisThrLeuPheCysValGlyArgCysPheLysTyrAspCysPheLeuHis----- 297  
DB 841 TTTTCATAGCTTTTCTCTTATGAGCGATGTTTAAATATATACGCTTCAATCGTAAATGC 900  
QY 298 -----ProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeu 315  
DB 901 AATTATTTCTTTATGAGCAACCCCAACATTTATAGCGGAAGAACACAAAGAACGCTCTA 960  
QY 316 AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluGluAlaLysGluPheAla 335  
DB 961 GACAAACAACCTTGTGAGACCAAGTGTTCACAGATTTGGAGGGAGCAAGAGATTGTCT 1020  
QY 336 AlaAlaLeuThrAlaGluArgGlnLeuValThrProProLysArgProGlyGlyAlaArgArg 355  
DB 1021 GCTGCTCTCAACGCTGAGCGGATTAAGACCCCAACCAACGTCACGAGAGCGCCGAGAGA 1080  
QY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375  
DB 1081 GAGCGGCTTCCCAATTAACATGACAGGCCCAAGCAACCCCAACATTAATGTCTGAAATCA 1140  
QY 376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGluAsnAspLys 395  
DB 1141 AAGGATACAGACAGTATTAAGGAAGCAGGAGCTGAACGGGGGAGAAACATATGAATAA 1200  
QY 396 GluGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415  
DB 1201 GAAGAAAGAGAAAGAAATGAACCTTCAGCTCTCTGAAGCAATTCCTGCTGTCAA 1260  
QY 416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAla 435  
DB 1261 ACACCAATTAAGATGAAGCCCAATATTTGAACCTCTAGAAATGTGAGATGAGTGCT 1320  
QY 436 GluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla 455  
DB 1321 GAACCTCAATGTTTAAAGTCTCTCATTTGGCACTTATGACAAATTTCTGTGCAATGCT 1380  
QY 456 ArgLeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer 475  
DB 1381 AGCTTAATTTGGACCAAAACATGTGAACAGGTGTATGATTTAAGTCAAAAGATCTAGC 1440  
QY 476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysValGlyLys 495  
DB 1441 ATCATAGCTCCAGCTCCCGCTGAGAGATGTGATCTCTCCAGAGAAAGAAAGAGAA 1500  
QY 496 HisArgLeuThrAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsn 515  
DB 1501 CACCGGTTGTGGGTGACACTGCAAGAAAGTACAGCTGAAGAAAGAGCGGCTCTTAAAC 1560  
QY 516 HisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysPro 535  
DB 1561 CATGTTTAACAATTCACACCTGTGATCATTCAGGAGCGCTTGTGACAGTTCGTGCTCT 1620  
QY 536 CysValIleIleAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsn 555  
DB 1621 TGTGTGATGACAAATTTTGTGAAAGTTTGTCAATGTAATCTCAAGTGTCAAAAC 1680  
QY 556 ArgPheProGlyCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeu 575  
DB 1681 CGCTTTCGGGATCCCGCTGCAAGCAAGTGAACCAAGCAGTCCCGCTCTACCTG 1740  
QY 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAsp 595  
DB 1741 GCTGTCCGAGGTGTGACCCGACCTCTGTTTACTGTGAGCGGCTGACCATTTGGAC 1800  
QY 596 SerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeu 615

DB 1801 AGTAAATATGTCCTGCAAGAACTGACATTCAGCGGGCTCCAAAAGCATCTAATG 1860  
 QY 616 LeuAlaProSerAspValAlaGlyTTPGylIlePheIleLysAspProValGlnLysAsn 635  
 DB 1861 CTGGACACCACTGACCGTGGACAGCTGGGGGATTTTATCAAGATCTCTGCAGAAAAT 1920  
 QY 636 GluPheIleSerGlyTyrCysGlyGluIleIleSerGlnAspGlnAlaAspArgArgGly 655  
 DB 1921 GAATTCATCTCGAATACCTGCGAGAGATTAATTTCTCAAGATGAACTGACAGAGAGG 1980  
 QY 656 LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheVal 675  
 DB 1981 AAAGGTATGATTAATAATACATGACGCTTCTGTTCACTTGACATGATATTTTGGGTG 2040  
 QY 676 AppAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695  
 DB 2041 GATGCAACCCGCAAGGGTAACAAAATTCCTTTGGCAAAATCATTCGGTAAATCAAACTGC 2100  
 QY 696 TyrAlaLysValMetMetValAsnGlyAspHisArgGlyIlePheAlaLysArgAla 715  
 DB 2101 TATGCAAAAGTTATGATGATGATTAACGGTATCAACAGATATGATTTTCCCAAGAGAGCC 2160  
 QY 716 IleGlnThrGlyGlnGluLeuPheAspTyrArgTyrSerGlnAlaAspAlaLeuLys 735  
 DB 2161 ATCCAGACTGGGAGAGAGCTGTTTTTGTATTAACATTAACAGCAAGCTGATGCCCTGAAG 2220  
 QY 736 TyrValGlyIleGlnArgGlnMetGluIlePro 746  
 DB 2221 TATGTCGCATCGAAAAGAAATGGAATCCCT 2253

## RESULT 15

ADP95922 standard; DNA; 2695 BP.  
 ID ADP95922 standard; DNA; 2695 BP.

AC ADP95922;

DT 07-OCT-2004 (first entry)

XX Human E2H2 (enhancer of zeste homologue 2) -encoding DNA, SEQ ID NO:3.

DE Human; E2H2; enhancer of zeste homologue 2; chromosome 7q35;  
 KW histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
 KW inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
 KW lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
 KW prophylaxis; prevention; diagnosis; cytostatic; gene; ds.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 167..2422

FT /tag= a /product= "E2H2 (enhancer of zeste homologue 2)"

PN W02004058969-A1.

PD 15-JUL-2004.

PR 22-DEC-2003; 2003WO-JP016417.

PR 24-DEC-2002; 2002JP-00373144.

PA (TAKE ) TAKEDA CHEM IND LTD.

XX Hikiichi Y, Nishizawa S;

DR WPI: 2004-525885/50.

DR P-PsDB; ADP95920.

PT Novel prophylactic and therapeutic agent of cancer, having compound or  
 its salt that inhibits activity of enhancer of zeste Drosophila homologue  
 protein, useful for treating cancer such as lung cancer, breast cancer.

PS Example 3; SEQ ID NO 3; 102pp; Japanese.

XX The invention relates to an agent for the prophylaxis and therapy of  
 CC cancer comprising a compound which inhibits the activity or expression of  
 CC human E2H2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
 CC E2H2 inhibitors include anti-E2H2 antibodies and E2H2 antisense nucleic  
 CC acids. The invention also relates to a kit for screening for inhibitors  
 CC of E2H2 activity or expression. E2H2 is a component of a complex with  
 CC histone methyltransferase activity which methylates histone H3 on lysine  
 CC residues at position 9 and/or 27. It has been found to be deregulated in  
 CC various tumours, and acts as a mediator of tumorigenesis, being able to  
 CC reverse repression of cyclin A which in turn mediates cell cycle  
 CC advancement. The agent of the invention can be used for preventing,  
 CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
 CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
 CC in drug screening. The agent of the invention is safe and less toxic than  
 CC prior art cancer prophylactic/therapeutic agents. The present sequence  
 CC represents a DNA encoding human E2H2 which was used in an example of the  
 CC invention.

SQ Sequence 2695 BP; 839 A; 549 C; 660 G; 647 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3,096-300 Length: 2695  
 Score: 4030.50 Matches: 745  
 Percent Similarity: 99.24 Conservative: 0  
 Best Local Similarity: 99.24 Mismatches: 1  
 Query Match: 99.54 Indels: 5  
 DB: 12 Gaps: 1

US-10-773-302-2 (1-746) x ADP95922 (1-2695)

QY 1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysThrArgLysArgValLys 20  
 DB 167 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTGTGGCGGAAGGCTGTAATA 226  
 QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
 DB 227 TCAGAGTCAATCCAGCTAGACAGCTCAAGAGCTTCAACAGCTGATGAAATTAAGACT 286  
 QY 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgGlnIleLeuAsnGlnGluTTP 60  
 DB 287 ATGTTAGTCCAAATCGTCAGAAATTTGGAAAGAACGAAATCTTAAACCAAGATGG 346  
 QY 61 LysGlnArgGlnIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
 DB 347 AAACAGCCAGAGATACAGCTGTGCAATCTGACCTTCGTGAGCTCATTTGGCGGAGCT 406  
 QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
 DB 407 ACGAGATGTGCGTACCAAGTACGACTTGATTTCCAAACAGATCATCCATTAAAGACT 466  
 QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTTPSerProLeuGlnAsnPhe 120  
 DB 467 CTGAATGACAGTTGCTTCACTACCAATATGATTTCTGTGTCCTCCCTACAGCAATTTT 526  
 QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
 DB 527 ATGTGGAAAGATGAAACGTTTTCATTAACATTCCTTATATGAGGAGAAAGATTTTAGAT 586  
 QY 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
 DB 587 CAGAGTGTACTTCTTATTAAGAACTAATAAATAATTAATGAGGAGAAAGTACACGGAGAT 646  
 QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGln 180  
 DB 647 AGAGATGTGGGTTATATAATGATGAATTTTGTGGAGTTGTGTAATGCCCTTGCTCA 706  
 QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGlnArgGlnGlnLysGln 200  
 DB 707 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
 QY 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220



Dh 767 AAAGATCTGAGATCAACGAGATGATAAGAAAGCCGCACTTCGGAATTTCTCTT 826  
Qy 221 Aspylsllephglnalaleseermetpheproaplysglythrallaglnleu 240  
Dh 827 GATAAATTTTGAAGCAATTTCTCAATGTTTCCAGTTAAGGACACAGAAAGAACTA 886  
Qy 241 Lysglnly 260  
Dh 887 AAGGAAAAATATAAGAACTACCGCAACAGAGCTCCAGGCGCACTTCTCTCAATGT 946  
Qy 261 ThrProenlleaprglyProasnalleyservalglnargglnuglnserleuhseser 280  
Dh 947 ACCCCCAATAGATGACCAATGCTAATCTGTTCAAGAGCAAAAGCTTACATCC 1006  
Qy 281 PhehlerleupheCyseArgCysePhelelylylylylylylylylylylylyly 297  
Dh 1007 TTTCAATGCTTTTCTGAGCGATGTTTAAATATGATGCTGCTTCAATCGTAAGTGC 1066  
Qy 298 -----ProPhehlsalathrProasnThrlylylylylylylylylylylyly 315  
Dh 1067 AATTATCTTTTCAATGCAACACCACTTATAGCGAAAGAACACAGAAACAGCTCTA 1126  
Qy 316 AsphenlyvsProsglyProgllyCysePhehlsleuglnuglnalalyeglnPhehla 335  
Dh 1127 GACACAAACCTTGAGACCAAGTGTACAGCATTTGAGAGGACAAAGAGATTGCT 1186  
Qy 336 Alaalaleuthralaglnargllylylylylylylylylylylylylylylylyly 355  
Dh 1187 GCTGCTCCACCGCTGAGCGATTAAGACCCCAACAAACGTCAGAGGCGCGAGAA 1246  
Qy 356 GlyargleuproasnenseerargProserThrProthrleasnvalleuglnser 375  
Dh 1247 GAGCGCTTCCCAATTAACAGTAGAGCGGCCAGACCCCACTTAATAGTCTGGAACA 1306  
Qy 376 LysasprThrapsereasprargglnalaglythrlylylylylylylylylylyly 395  
Dh 1307 AAGGATACAGACAGTAGAGAGACAGGACTGAACGGGGGAGAAACATGATAA 1366  
Qy 396 Glnuglnuglnly 415  
Dh 1367 GAAG 1426  
Qy 416 ThrProilleylmetlyProasnleuglnProgluasnvalglutPserglyala 435  
Dh 1427 ACACCAATTAAGATGAGACCAATTAATGACCTCTGAGAAATGTGAGTAGGTGCT 1486  
Qy 436 Glnlasermetpheargvalleuilegilythrlylylylylylylylylylylyly 455  
Dh 1487 GAAGCTCAATGTTTGAAGTCTCTATGCACTTACTATGACATTTCTGTGCCATTCCT 1546  
Qy 456 Argleuilegilythrly 475  
Dh 1547 AGGTTAATTTGGGACCAAAACATGTAGACAGGTATGAGTTTAAAGTCAAAATCTAGC 1606  
Qy 476 IlellealaproalaproalaglnasprvalasprThrProarglylylylylyly 495  
Dh 1607 ATCATAGCTCCAGCTCCGCTGAGATGTGATATCTCTCCAGGAAAAAGAGGAAA 1666  
Qy 496 HisargleuThrallalaleCysearglylylylylylylylylylylylylylyly 515  
Dh 1667 CACCGGTGTGGGCTGACACGACAGAAAGATACGCTGAAAGAGACGCGCTCTTAC 1726  
Qy 516 Hisvally 535  
Dh 1727 CATGTTTCAACTATCAACCTGTGATCATCCAGCGAGCTTGTGACAGTTCTGTGCT 1786  
Qy 536 CysvallealeaglnasprheCysegllylylylylylylylylylylylylyly 555  
Dh 1787 TGTGTATAGACCAAAATTTTGTGAAAAAGTTTGTCAATGATGTTCAAGAGTCAAAAC 1846  
Qy 556 ArgpheProglyCyseargCysevalaglnCyseasnThrlylylylylylylylyly 575  
Dh 1847 CGCTTCCGGAGTGCCTGCAAGACAGTGCAACACCAAGACAGTGCCTGTGCTGCTG 1906

Qy 576 Alavalarglnly 595  
Dh 1907 GCTGTCCGAGAGTGTGACCTGTGACCTGTGTCTTACTGTGTGAGCGCGTGAACATTTGGGAC 1966  
Qy 596 SerlybasenvalserCyselybasenCyselelleaglnargglyserlylylylyly 615  
Dh 1967 AGTAAATATGTCTTCAAGAACTGCAATTCAGCGGGGCTCCAAAGACATCTATTG 2026  
Qy 616 Leuallaprosereaspralalaglylylylylylylylylylylylylylylylyly 635  
Dh 2027 CTGGCAACATCTGACGATGCGAGCGCTGGGGATTTTTCAAAAGATCTGTGCAAGAAAAT 2086  
Qy 636 Glnupheleesergly 655  
Dh 2087 GAATTCATCTCAGAAATCTGTGAGAGATTAATTTCTCAAGATGAACTGACAGAAAGGG 2146  
Qy 656 Lysvally 675  
Dh 2147 AAAGTATGATTAATATCATGTGCACTTCTGTCAACTGAACATGATTTTGTGCTG 2206  
Qy 676 Aspalathrargly 695  
Dh 2207 GATCAACCGGCAAGGTTAACAAATTCGTTTGCAATCATTCGTAATCCAACTGC 2266  
Qy 696 Tyrallalyvalmetmetvalasnglyasprhisargllelylylylylylylyly 715  
Dh 2267 TATCAAAAGTTATGATGTGTTAACGTGTATCACAGATGATATTTTGTCCAGAGAGCC 2326  
Qy 716 Ilelgnhrly 735  
Dh 2327 ATCCAGACTGCGAAGAGCTGTTTGTGATTAACAGATACAGCCAGGCTGATGCCCTGAG 2386  
Qy 736 TyrvalglylleaglnargglnumetglnlePro 746  
Dh 2387 TATGTGCGCATCGAAAGAAATGAAATCCCT 2419

Search completed: August 13, 2006, 13:49:22  
Job time : 1222.01 secs

[illegible]

Kono, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system-384-format  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Atakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Pleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotman, M., Hume, D.A., Kamya, M., Lee, N.H., Lyons, P., Rong, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L., Wyshaw-Borle, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohzuki, S. and Hayashizaki, Y.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

5 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chotaka, C., Corbani, L.B., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G., Petrovsky, N., Pili, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempole, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Borle, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirokawa, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imochi, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)

6 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lennard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Bremner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, V., Allen, J.E.,

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)

7 Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Matlick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lennard, B. and Wahlestedt, C.

Riken Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)

8 (bases 1 to 2652)

Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Horii, P., Iida, Y., Imanura, K., Imochi, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Kono, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa, 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Hiromitsu Nakaguchi (Center for Experimental Medicine, Institute of Medical Science, University of Tokyo 4-6-1, Shirokane-dai, Minato-ku, Tokyo, 108-8639, JAPAN) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://lancm.gsc.riken.jp/>

## FEATURES

source

Location/Qualifiers

1..2652

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM\_DB:K0C0026809"

/db\_xref="taxon:10090"

/clone="K0C0026809"

/cell\_type="erythroblast"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

146..2386

/note="unmated protein product; enhancer of zeste homolog

2 (Drosophila) (MGD|WGI:107940 GB|NM\_007971, evidence:

BLASTN, 99%, match=2635)

putative"

/protein\_id="BAB34232.1"

/db\_xref="GI:74186126"

/translation="MGQTGKSEKGPVCRRKRVKSEYRLNQLKFRPADVKTMESS

NRKILERTETLNDQKQRIQHVIMTSVSLRGTRCSTVSDLPDPAQVPLKTLN

AAVAVIMYSPLQDNFMEDETVLHNIPLYGDEVLDQDGFIELIKNDGKVGHD

RBCGRINDEIFVELYNALGOYNDDDDDGDDPDRERKQKQLEJNRDCKECPKRF

PADKIREAISMPDKGTAEELKEXKELTQQLRGALPRECTPIIDGPNKSVRBO

SLHSFHTLPFCRCFKYDCFLPHFATPTTKNTETLNDNCPGPOCTQHDGAKER

AAALTAERIKTPPKRPGRRGRPLPNSSRSTPTISLVESKDTSDREAGETEGEN

## Alignment Scores:

Pred. No.: 0 Length: 2652  
Percent Similarity: 99.00 Matches: 733  
Best Local Similarity: 99.2% Conservative: 7  
Query Match: 98.3% Mismatches: 6  
DB: 6 Gaps: 0

US-10-773-302-2 (1-746) x AKI57853 (1-2652)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB 146 ATGGGCGACAGCTGGGAGAAATCTGAGAAAGGACCGGTTGTGGCGGAAAGCTGTAA 205  
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgGlnLysArgLysSer 40  
DB 206 TCAGAGTCATGAGCTGAGACAGCTCAAGAGTTCAGAAAGCTGATGAAAGTAAAGCT 265  
QY 41 MetPheSerSerAspArgGlnLysIleLeuGluArgThrGlnIleLeuAsnGlnGluTrp 60  
DB 266 ATGTTTAAGTTCATCGTCAGAAATTTTGGAAAGAACTGAAACCTTAAACAGAGATGG 325  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgLysThr 80  
DB 326 AAGGACGGGAGGATACAGATGTCATCATCATCTTCGTGAGCTCATTTGGCGGAGCT 385  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 386 AAGGAGGTTCAGTACCAAGTACCTTGATTTTCCAGCAAGATCATCCGTTAAAGACC 445  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnIle 120  
DB 446 CTGAATGACGTCGCTCGGTGCTTAATGTAATCTTGTGCGCCCTTACACAGAAATTTT 505  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 506 ATGGGGAAGAGCAAGAACTGTTTACATTAACATTCCTTAATGGGGAGAGAGTTCTGGAT 565  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisIleLysAsp 160  
DB 566 CAGATGGACACTTATTAAGAACTAAATAAATTTATGATGAGAAAGTGCATGATGAC 625  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnGln 180

DB 626 AGAGAAATGAGATTATTAATATGATGAATTTTGTGAGTTGGTAAATGCTCTTGATCA 685  
QY 181 TyrAsnAspAspAspAspAspAspGlyLysAspProGluGluArgGluGluLysGln 200  
DB 686 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 745  
QY 201 LysAspLeuGluAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 746 AAAGATCTAGAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 805  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 806 GATAAATATATTGAAGCCATTCTCTCAATGTTTCCAGATTAAGGACCGCAGAAAGACTG 865  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGluLys 260  
DB 866 AAAGAAAATATTAAGAACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 926 ACTCCAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 985  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisPhePheHis 300  
DB 986 TTTGATACGCTCTTCTGTGCAAGATGTTTAAGTATGATGATGATGATGATGATGATGAT 1045  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 1046 GCAACACCCCAACACATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105  
QY 321 GlyProGlnCysTyrGlnHisLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1106 GATCCACAGTGTATCCAGCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1165  
QY 341 GlnArgIleLysThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360  
DB 1166 GAGGCTATTAAGACACCACTTAACGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1225  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1226 AACGATGACAGACCCAGACACCCCAACATCTGAGTGTGAGTCAAAAGATACACAGCT 1285  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyLysAsnAspLysGluGluGluLys 400  
DB 1286 GACAG 1345  
QY 401 LysAspGluThrSerSerSerSerSerGlnIleAsnSerArgCysGlnThrProIleLysMet 420  
DB 1346 AAAGATGAGAGCTCAGCTCTCTGAGCAAAATCTGGGTGTCAAAAGCAATTAAGATG 1405  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1406 AAGCCAAATATTAAGACCTCTGAGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1465  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnIleCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1466 AAGGCTCTCATTTGATCTTACGATTAATCTTTGTGCAATTTGAGGCTTAATTTGGAGCC 1525  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB 1526 AAAACATGTATACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1585  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
DB 1586 CCCACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1645  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspLysSerSerAsnHisValTyrAspTyr 520  
DB 1646 GACATCTGCAAGAAAGATACAACTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1706 CAACCTGTGACATTCACGAG 1765

Qy	541	AanhheCySgIuLyPhePheCySgIuCySgSerSerGluCySgIuPanaMaPheProGluCyS	560
Db	1766	AATTtTGTGAAAAGTTTGTCTCAAGTGTCAAGTGTCAAAACCGTTTCTCGAATGT	1825
Qy	561	ArgCySlySylagIuCySaSnThxLySgIuCySProCySlyTLeuAlaValArgGluCyS	580
Db	1826	CGGTGCAAAGCACAATGCAACGCCAAACAGTGTCCATCTCACTGTGTCCGAGAGTGT	1885
Qy	581	AspProAspLeuCySleuthThCySgIuAlaAlaAspHisTTPaPaspSerLyAsnValSer	600
Db	1886	GACCCCTGACCTGTGTCTCAAGTGTGAGACTGTCAACATTGGAGACAGTAAAAATGTATCC	1945
Qy	601	CysLySaSnCySserTLeGluArgGlySerLySlySHsleuLeuLeuAlaProSerAsp	620
Db	1946	TGTAAAGACTGTAGTACATTCAAGCGGGGCTCTTAAAAGCACTTACTGTGGACCGCTGTAT	2005
Qy	621	ValAlaGlyTTPGlyTLePheTLeLySaAspProValGluLySaSnGluPheTLeSerGlu	640
Db	2006	GTCGACAGCTGGGGCATCTTTATCAATCAAGATCCGTGACAGAAATGAATTCATCTCAGAA	2065
Qy	641	TyrCySgIyGluTLeTLeSerGluAspGluAlaAspAArgGlyLySValTyrAspLyS	660
Db	2066	TACTGTGGGAGATTATTCTCAGGATAGAAGACAGAAAGAGAAAGTGTATGACAA	2125
Qy	661	TyrMetCySserPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLyS	680
Db	2126	TACATGTGCACACTTCTGTGTCACTGTGAACAATGATTTTGTGTGTGATGACAAACCGAAAG	2185
Qy	681	GlyAsnLySITLeArgPheAlaAsnHisSerValaAspProAsnCySlyTAlaLySValMet	700
Db	2186	GGCACAAATAATTCGTTTGTCTTATCATTCATTCAGTAATCCAAATCTGCTAATGCAAAAGTTATG	2245
Qy	701	MetValaAsnGlyAspHisArgGlyTLeGlyTLePheAlaLySaArgAlaTLeGluThGlyGlu	720
Db	2246	ATGCTTAATGGTGACACAGATAGGACATCTTGTCTAAAGGGGCTATCAAGCTGGTAA	2305
Qy	721	GluLeuPhePheAspTyrArgTyrSerGluAlaAspAlaLeuLySlyTyrValGlyTLeGlu	740
Db	2306	GAGTTGTTTTTTTGATTAACAGTACACGCCAGGCTGATGCCCTGAATATGTGGGCATCGAA	2365
Qy	741	ArgGluMetGluTLePro	746
Db	2366	CGAGAAATGCAAAATCCCT	2383
RESULT 2			
AY39441			
LOCUS	AY39441	2241 bp	DNA linear
DEFINITION	Homo sapiens EZH2 gene, VIRUTUAL TRANSCRIPT, partial sequence,		Genomic survey sequence.
ACCESSION	AY39441		
VERSION	AY39441.1	GI:39755430	
KEYWORDS	GSS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini;		
	Homnidae; Homo.		
	1 (bases 1 to 2241)		
	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
	Todd,M.A., Tanendau,D.M., Civello,D.R., Lu,F., Murphy,B.,		
	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,		
	Adams,M.D. and Cargill,M.		
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous		
	gene trices		
JOURNAL	Science	302 (5652),	1960-1963 (2003)
PUBMED	14671302		
REFERENCE	2 (bases 1 to 2241)		
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
	Todd,M.A., Tanendau,D.M., Civello,D.R., Lu,F., Murphy,B.,		
	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,		
	Adams,M.D. and Cargill,M.		
TITLE	Direct Submission		

JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES	Location/Qualifiers
source	1..2241
gene	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" <1..>2241 /gene="EZH2" /locus_tag="HGM0225"
ORIGIN	
Alignment Scores:	
Pred. No.:	3.55e-291
Score:	3250.00
Percent Similarity:	83.1%
Best Local Similarity:	83.1%
Query Match:	80.2%
DB:	14 Gaps: 0
US-10-773-302-2 (1-746) x AY399441 (1-2241)	
QY	1 MetGlgIntHnrgILyVAlYsSerGILySgILyProValCYeTTPArGlySArGVallys 20
Db	1 ATGGGCGACACTGGAGAGAAATCTAGAGAGGACCAAGTTTGTGGCGGAAGCGTGAAGA 60
QY	21 SerGILyMetArGLyLeuArGLyLeuLySArPheArGAlaApGLyVallySer 40
Db	61 TCAGAGTACATGCCACTGACGACGCTCAAGAGTTCAAGCGAGCTGATGAATGAAGAGT 120
QY	41 MetPheSerSerAenArGLyLysIleLeuGILyArGThrGILyIleLeuAnGInGILyTP 60
Db	121 ATGTTTAACTCCAAATCGTCAGAAATTTTGGAAAGACGGAATCTTAAACCAAGATG 180
QY	61 LysGILyAGArGILyGlnProValHisIleLeuThrSerValSerSerLeuArGLyThr 80
Db	181 AAACGCGAAGATACAGCTGTGACATCTGACTCTGTAGGCTCATTTGGCGGGAGCT 240
QY	81 ArgGILySerValThrSerAspLeuAspPheProThrGILyValIleProLeuLySThr 100
Db	241 AGCGAGGNN 300
QY	101 LeuAenAlaValAlaSerValProIleMetYrSerTTPSerProLeuGInGInAspPhe 120
Db	301 NNN 360
QY	121 MetValGILyAspGILyThrValLeuHisAenIleProYrMetGILyAspGILyValLeuAsp 140
Db	361 NNNNGGAGAGGAAACTGTTTACATTAACATCTCTATATGGAGATGAAGTTTAGAT 420
QY	141 GlnAspGILyThrPheIleGILyLeuLeuIleYsAsnYrAspGILyValHisGILyAsp 160
Db	421 CAGGATGTGATCTTTCATTGAAGAACTAATAAATTAAGATGGGAAGTACCGGGAT 480
QY	161 ArgGILyCgILyPheIleAsnAspGILyIlePheValGILyLeuValAsnAlaLeuGILyGln 180
Db	481 AGAGATGTGGGTATTAATATGATAAATTTTGTGAGAGTGTGAATCCCTTGGTCAA 540
QY	181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200
Db	541 TATAATGATGATGACGATGATATATATGAGACGATCTGAAAGAAAGAGAAAGAAAGCAG 600
QY	201 LysAspLeuGILyAspHisArgAspAspLySgILySerArgProProArgILySPhenProSer 220
Db	601 AAAGATCTGGAGGATCACCGAGATGATTAAGAAAGCCGCCACCTGGAAATTTCTTCT 660
QY	221 AspLySILyPheGILyAlaIleSerSerMetPheProAspLySgILyThrAlaGILyLeu 240
Db	661 GATTAATATTTTGAAGCACTTCTCTCAATGTTTCCAGATAGGGGACACAGAAAGAACTA 720
QY	241 LysGILyLySILyLeuThrGILyGInGILyLeuProGILyAlaLeuProProGILyCys 260

[illegible]

Db	1801	TGCAAGAACTGCAGATATTACAGCGGGGCTCAGAAAAGATCATATTGCTGGACCATCTGCAC	1800
Qy	621	ValAlaGlyTTPGIYIlePheIleLysAspProValGlnLysaGlnPheIleSerGlu	640
Db	1861	GTGCACAGGCTGGGGGATTTTATTCAGAAATCTCTGTGCAGAAAATGAAATTCATCTCAGAA	1920
Qy	641	TyrCysGlyYguIleIleSerGlnAspGlnAlaAspArgArgGlyLysValTyrAspLys	660
Db	1921	TACTGTGAGAGAAATTAATTTCTCAAGATGAAAGCTGACAGAAAGGAAAGGTATGATGAA	1980
Qy	661	TyMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys	680
Db	1961	TACATGTGCAGCTTCTTGTTCAATCTTGACATGANNNNNNNNNNNNNNNNNNNNNNNNNN	2040
Qy	681	GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet	700
Db	2041	NN	2100
Qy	701	MetValaGlnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu	720
Db	2101	NN	2160
Qy	721	GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu	740
Db	2161	NN	2220
Qy	741	ArgGluMetGluIlePro 746	
Db	2221	AGAGAAATCGAAATCCCT 2238	
RESULT 3			
AY399443		2241 bp	DNA
LOCUS			linear
DEFINITION	Mus musculus EHz2 gene, VIRUTAL TRANSCRIPT, partial sequence,	GSS 15-DEC-2003	
ACCESSION	AY399443		
VERSION	AY399443.1		
KEYWORDS	GSS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
	Sciurognathu; Murioidea; Muridae; Mus.		
REFERENCE			
AUTHORS	1 (bases 1 to 2241)		
	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
	Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,		
	Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,		
	Adams,M.D. and Cargill,M.		
TITLE	Interfering nonneutral evolution from human-chimp-mouse orthologous		
	gene trios		
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBMED	14671302		
REFERENCES	2 (bases 1 to 2241)		
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
	Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,		
	Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,		
	Adams,M.D. and Cargill,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,		
	Rockville, MD 20850, USA		
COMMENT	This sequence was made by sequencing genomic exons and ordering		
	them based on alignment.		
FEATURES	Location/Qualifiers		
source	1..2241		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
gene	/db_xref="taxon:10090"		
	<1..2241		
	/gene="EHz2"		
	/locus_tag="HGM0225"		
ORIGIN			
Alignment Scores:			

Pred. No.: 8,49e-287 Length: 2241  
Score: 3203.00 Matches: 609  
Percent Similarity: 82.6% Conservative: 7  
Best Local Similarity: 81.6% Mismatches: 130  
Query Match: 79.1% Indels: 0  
DB: 14 Gaps: 0

US-10-773-302-2 (1-746) x AY399443 (1-2241)

QY 1 MetGlyGlnThrGlyLysLeuSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB 1 ATGGGCCGACTGGGAGAAATCTGAGAAAGGACCGGTTGTGGCGGAACGGTATAA 60  
QY 21 SerGluTrpMetArgLeuArgGlnLeuLysArgPheArgArgAlaSerGluValLysSer 40  
DB 61 TCAGAGTCATGAGACTGAGACAGCTCAAGAGTTCAAGAGCTGATGAAGTAAAGACT 120  
QY 41 MetPheSerSerAspAspArgGlnLysLleuGluLysArgGluLleuLysGlnGluTrp 60  
DB 121 ATGTTAGTCCATCGTCAGAAATTTTGAAAGAACCTGAACCTTAAACCAAGAGTGG 180  
QY 61 LysGlnArgArgLleGlnProValHisLleuThrSerValSerSerLeuArgGlyThr 80  
DB 181 AAGCAGCGGAGATACAGCCTGTGCACATCAGACTTCTGAGCTCATTCGCGGAGACT 240  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValLleProLeuLysThr 100  
DB 241 AGGAGGNN 300  
QY 101 LeuAsnAlaValAlaSerValProLleMetYrsTrpSerProLeuGlnLysAspHe 120  
DB 301 NNN 360  
QY 121 MetValGluAspGluThrValLleuHisAsnLleProYrMetGlyAspGluValLleuAsp 140  
DB 361 NNNGTGAAAGCAAACTGTTTACATTAACATTCCTTATATGGGGAGTGAAGTCTGAT 420  
QY 141 GlnAspGlyThrPheLleGluLleuLleLysAsnYrAspGlyLysValHisGlyAsp 160  
DB 421 CAGAGTGGCACTTTCATTGAAGAACTTAATAAAATATGATGAAAGTGCATGTGATC 480  
QY 161 ArgGluCysGlyPheLleAsnAspGluLlePheValGluLeuValAsnAlleuGlyGln 180  
DB 481 AGAGATGTGATTTATAAATGATGAAATTTTGTGAGTGGTAAAGCTTGTGCTAA 540  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGluLysGln 200  
DB 541 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 601 AAAGATCTCAAGGATATCGAGATGATTAAGAACTTGCCCACTCGGAAATTTCTGCT 660  
QY 221 AspLysLlePheGlnAlleLleSerSerMetPheProAspLysGlyThrAlaGluLleu 240  
DB 661 GATAAAATATTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGCGACCGCAGAAACATG 720  
QY 241 LysGluLysTrpLysGluLleuThrGluGlnGlnLysProGlyAlaLleuProProGluCys 260  
DB 721 AAAGAAATATTAAGAACTCAGGAGCGAGCTCCAGGCTGCTGCTGCTGCTGCTGCT 780  
QY 261 ThrProAsnLleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLleuHisSer 280  
DB 781 ACTCCAAACATCGATGAGCAAAATGCAAAATCTGTTCAAGAGGAGCAAAAGCTTGCAATCA 840  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLleuHisProPheHis 300  
DB 841 TTTCTATAGCTCTTCTGTGCGAGATGTTTAAAGTATGACTGCTTCTCATCTNNNNNNNN 900  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLleuAspAsnLysProCys 320  
DB 901 NNN 960

QY 321 GlnProGlnCysTrpGlnHisLleuGlnGlyValLysGluLysPheAlaAlaLleuThrAla 340  
DB 961 NNN 1020  
QY 341 GlnArgLleLysThrProProLysArgProGlyGlyYrArgArgArgGlyArgLeuProAsn 360  
DB 1021 GAGCTTAAAGACACCACTTAAGCCCGGAGGCGCGCAGAAAGAGAAAGACTTCCGAT 1080  
QY 361 AsnSerSerArgProSerThrProThrLleAsnValleuGluSerLysAspThrAspSer 380  
DB 1081 AACAGTAGAGACCCCAACCCCAACATCATGTGCTGAGTCAAGAGTATACAGACAGT 1140  
QY 381 AspArgGlnAlaGlyThrGluThrGluGlyGlnAspAspAspLysGluGlnGluLys 400  
DB 1141 GACAGAAAGCAGAGCTGAACTGGGAGAGAAACATATTAAGAAAGAAAGAGAA 1200  
QY 401 LysAspGluThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProLleLysMet 420  
DB 1201 AAAGATAGAGCTCAAGCTCTCTGAAGCAAAATTCGGGTCTCAACACCAATTAAGATG 1260  
QY 421 LysProAsnLleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1261 AAGCCAATATATGAACCTCCTGAGAAATGTGAGTGGAGTGGTGTGAAGCTTCATGTTT 1320  
QY 441 ArgValLleuLleGlyThrTyrTrpAspAspPheCysAlaLleAlaArgLleuLleGlyThr 460  
DB 1321 AGAGTCTCATGTGTACTACTAGTAACTTAATCTTTGTCCATGTCTAGGCTTAATGGAGCC 1380  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerLleLleAlaProAla 480  
DB 1381 AAACATATGACAGCGGTATGAGTTTAGTCAAGAGTCAATATCATAGACACCTGTT 1440  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
DB 1441 CCCACTGAGATGTAACACTCCTCAAGAAAGTGAAGTGGTGTGGGCT 1500  
QY 501 AlaHisCysAspArgLleGlnLleuLysLysAspGlySerSerAsnHisValTyrAsnTrp 520  
DB 1501 GCACATGCAAAAGATACAACTAAAGAGCGGCTCTTAAACATGTTTACAACTAT 1560  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValLleAlaGln 540  
DB 1561 CAACTCTGTACATCCAGCGGACGCTTGTGACAGTTCGTGCTGCTGTGATGACCAA 1620  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGlyCys 560  
DB 1621 AATTTTGTAAAGTTTGTCAATGTATGATCAAGTGTCAAAACCGCTTCTCGAGTGT 1680  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLysAlaValArgGluCys 580  
DB 1681 CGGTGCAAAAGCAGATGCAACCAACAGTGTCAATGCTACCTGGCTGTCCAGAGTGT 1740  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1741 GACCTTCACTCTGTCTCACTGTGAGCTCTCACTTGGAGCAGTAAATATGATACC 1800  
QY 601 CysLysAsnCysSerLleGlnArgLysTrpLysHisLleuLleuAlaProSerAsp 620  
DB 1801 TGTAAAGACTATGACATTCAGCGGGGCTTAAGAGCACTTAAGCTGGACCGTCTGAT 1860  
QY 621 ValAlaGlyTrpGlyLlePheLleLysAspProValGlnLysAsnGluLysLleSerGlu 640  
DB 1861 GTGCGAGGCTGGGCACTTTATCAAGATCTGTACGAAATTAATATATCTCAGAA 1920  
QY 641 TyrCysGluGluLleLleSerGlnAspGlyAlaAspArgArgGlyLysValTyrAspLys 660  
DB 1921 TACTGTGGGAGATATTTCTCAGATGAACCAACAGAAAGAGAAAGTATAGCAAA 1980  
QY 661 TyrMetCysSerPheLeuPheAsnLleuAsnAspPheValLleAspAlaThrArgLys 680  
DB 1981 TACATGTGCACTTCTGTCTCACTTGAACAATGNNNNNNNNNNNNNNNNNNNNNNNN 2040  
QY 681 GlnAsnLysLleArgPheAlaAsnHisSerValAsnProAsnCysTrpAlaLysValMet 700



[illegible][illegible]

QY 401 LysAspGluThrSerSerSerSerGlnAlaAsnSerArgGlnThrProIleuMet 420  
 DB 1201 AAAAGTGAACCTTGAGCTCTCTGAAACAATCTCGGTGCAAAACCAATGAAGATG 1260  
 QY 421 LysProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 DB 1261 AAGCAATATATGAACTCTGAGAAATGAGAGTGAGTGTGCTGAAGCCCTCAATGTTT 1320  
 QY 441 ArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 DB 1321 AGAGTCTCATTTGGCACTTACTACTGACAAATTTCTGTGCACTTGTGCTTAAGTTAATGGACC 1380  
 QY 461 LysThrCysArgGlnValTyrGluPheArgValIleGlyGlnSerSerIleIleAlaProAla 480  
 DB 1381 AAAACATGTAGACAGGTGTATGAGTTTGTAGTCAAAAGATCATGATCATGCTCCAGCT 1440  
 QY 481 ProAlaGluAspValAspThrProProArgTyrGlySerArgGlySerIleAlaGluThrAla 500  
 DB 1441 CCCGCTGAGATGTGGATATCTCTCCAGAGAAAAGAGAGAAACACCGGTTGTGGGCT 1500  
 QY 501 AlaHisCysArgGlyIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 DB 1501 GCAACTGCAAAAGATACAGCTGAAAAGAGACGCTCTCTTAACCATGTTTACACATAT 1560  
 QY 521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 DB 1561 CAACCTGTGATCATCCACGCGAGCTGTGACAGTTGCTGTCCTTGTGTATATGACAA 1620  
 QY 541 AsnPheCysGlyLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
 DB 1621 AATTTTGTGAAAGTTTGTCAAGTGTGTCAAGTATCAAAACCGCTTCCGGAGTGC 1680  
 QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
 DB 1681 CGCTGCAAGACAGTGCACACCAACAGAGTCCCTGCTCACTGCTGCTGAGAGTGT 1740  
 QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTyrPheSerIleAsnValSer 600  
 DB 1741 GACCTGACCTCTGCTTACTTGTGAGCCGCTGACCATTTGGGACAGATGAATAATGTGCC 1800  
 QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysIleuLeuLeuAlaProSerArg 620  
 DB 1801 TGCAGAACTGCAGATTCAGGGGGCTCCAAAAGCATCTATGCTGACACCATCTGAC 1860  
 QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 DB 1861 GTGGCAGGCTGGGGATTTTATTCAAAGATCTCTGCAAGAAAAGATGATTCATCTCAGA 1920  
 QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 DB 1921 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGCACAGAGGAGAAAGTGTATATATAA 1980  
 QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
 DB 1981 TACAAGTGCAGCTTCTGTCTCAACTGAACTGAACTGANNNNNNNNNNNNNNNNNNNNNN 2040  
 QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
 DB 2041 NNN 2100  
 QY 701 MetValaAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
 DB 2101 NNN 2160  
 QY 721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLysLysTyrValGlyIleGlu 740  
 DB 2161 NNN 2220  
 QY 741 ArgGluMetGluIlePro 746  
 DB 2221 AGAGAAATGGAATTCCT 2238

RESULT 5  
 AK154565  
 LOCUS  
 DEFINITION  
 Mus musculus MOD-derived CD1ic +ve dendritic cells cDNA, RIKEN full-length enriched library, clone: F630047M06 product: enhancer of zeste homolog 1 (Drosophila), full insert sequence.  
 ACCESSION  
 AK154565  
 VERSION  
 AK154565.1 GI:74181971  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 REFERENCE  
 1  
 Carninci, P. and Hayashizaki, Y.  
 TITLE  
 High-efficiency full-length cDNA cloning  
 JOURNAL  
 Meth. Enzymol. 303, 19-44 (1999)  
 PUBMED  
 10349636  
 REFERENCE  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 TITLE  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL  
 Genome Res. 10 (10), 1617-1630 (2000)  
 PUBMED  
 11042159  
 REFERENCE  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, D., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 TITLE  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL  
 Genome Res. 10 (11), 1757-1771 (2000)  
 PUBMED  
 11076861  
 REFERENCE  
 4  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nishikido, I., Peeble, G., Quackenbush, J., Schriml, L.M., Steubli, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamita, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaur, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilmink, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kontecki, S., and Hayashizaki, Y.  
 TITLE  
 RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium  
 JOURNAL  
 Functional annotation of a full-length mouse cDNA collection  
 NATURE 409 (6821), 685-690 (2001)  
 PUBMED  
 11217851  
 REFERENCE  
 5  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishikido, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brack, D., Brusk, V., Chochia, C., Corbani, L.B., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S.,

Guernich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziereki, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Malat, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perle, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Q. D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Veierd, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wymshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirokawa-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imochi, K., Itoh, Y., Itoh, M., Kagawa, T., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

CONSRITM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)

1246851

6  
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Bremner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aldins, V., Allen, J. E., Ambesi-Impombato, A., Anselotti, R., Arakawa, T., Arai, Y., Balleu, T. L., Banasi, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crome, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., Di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. P., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgi-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kobayashi, G., Kristiansen, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liu, S., Liu, S., McWilliam, S., Madan, B., Mader, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, P., Miyake, S., Morita, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Noji, P., Ohara, O., Okazaki, Y., Ohtsuka, V., Pang, K. C., Pavan, W. J., Pavoni, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Ros, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Sempke, C. A., Seno, S., Sees, L., Sheng, Y., Shibata, Y., Shimada, K., Shimada, K., Silva, D., Sinclaire, B., Sperling, S., Stupka, B., Sugita, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, B., Vazdarov, R., Wei, C. L., Yang, K., Yamashita, H., Zaborov, Y. E., Zhu, S., Zimmer, A., Hilde, W., Bull, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, K., Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Kono, H., Nakano, K., Nimomura, N., Nishio, T., Okada, M., Pleassey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanabe, K., Okamura-Ohno, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.

CONSRITM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

The transcriptome landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)

16141072

7  
Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yab, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J. S., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Peshkin, M. A., Sandelin, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlestedt, C.

CONSRITM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

RIKEN Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)

16141073

8 (bases 1 to 4089)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imochi, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Kono, H., Murata, M., Nakamura, M., Nimomura, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, K., Watanabe, M., and Hayashizaki, Y.

TITLE  
JOURNAL

Submitted (30-Mar-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

FEATURES

source

CDS

1. 4089  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="FANTOM DB: F630047M06"  
/db\_xref="taxon:10090"  
/clone\_type="NOD-derived CD11c +ve dendritic cells"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
92. 2344  
/note="unmated protein product; enhancer of zeste homolog 1 (Drosophila) (MGD|MGIT1097695|GB|U60453, evidence: BLASTN, 99%, match=3459)  
putative"  
/codon\_start=1  
/protein\_id="BAE32680.1"  
/db\_xref="GI:74181972"  
/translation="MRKMDIASPTSKITTYKRYKSEYKRLQKRLQNMKAKAL  
YVAPFAVQETQILNBERKRLVQVQPMKPVSGHPLKCALESTPFGDSODMLM  
RSLNVALVPIPTMYSPLQDNFVDEIVLQNIPTMGDEVEBETPEBILINNYDK  
VRHIEEMPGSVLLSDAVFLVNLQNYSDREBGNHDPDGKDDSKEDLPVPRK  
KRAIEGKSSKQPPNDMI FSAIYNSFPNGVDDMKRYRELTEKSDNALPPOC  
TPNIDGPAKSVQGEQSLHSRTYFCRCFCYDCTDTPHATPVYKXKEIKTEPE  
PCGTDFULLEGARVATLHPKSKSGRRRRRHPVVSASCSASAMATKESDSD

Alignment Scores:

Pred. No.:	Length:	4089
Score:	2606.50	492
Percent Similarity:	77.6%	96
Best Local Similarity:	64.9%	117
Query Match:	64.3%	53
DB:	6	13

US-10-773-302-2 (1-746) x AKU54565 (1-4089)

QY 15 TTPAryGaagvallysSerGluTyMeckArgLeuArgGlnLeuLyArgPheArgArg 34  
DB 143 TGGAAAGAAAGGAAATCGAGTATATGCGCTTCACAGCTCAACGCTCCAGCA 202  
QY 35 ALaapGvUaVlysserMetPheSerSeranArgGlnLySileLeuGluArgThrGlu 54

Db 203 AATATGGAGCAAGGCTGTGTATGTGGCAAAATTTTGCAAAAGTTCAAGAAAACCCAA 262  
Qy 55 ILeuubangIngluTPuLyEgInaArgIleGlInProValIhIleLeuThSerVal 74  
Db 263 ATCTTCATGAAAGGTGAAAGAACTTCGTCCAGCCTGTTCAAGCCATGAACCCGTG 322  
Qy 75 SerSerLeuArgIlyThrArgIuCySerValThrSerAepLeu--AepPheProThr 93  
Db 323 AGTGGGCAACCTTTTGCAAAAGGTGTGCTAGAGCATTTTCCAGGCTGACAGC 382  
Qy 94 GluValIleProLeuLyThrLeuAmaIaValaIaSerValProIleMetTyrSerThr 113  
Db 383 CAGATATGTGAATGGGTCTCTGAACCTGTGCACTGGTTCCTCATATATATTCGG 442  
Qy 114 SerProLeuGInglInAenPheMetValGluAepGluThrValIleuHleAenIleProTyr 133  
Db 443 TCCCTCTCCAGCAGAAATTCATGATGAGATGAGACGGTTTGTGCAATATTCATAC 502  
Qy 134 MetGlyAepGluValIleuAepGluAepGlyThrPheIleGluGluLeuIleLyAenTyr 153  
Db 503 ATGGGTGACGAGTGAAGAGAAAGATGAGACTTTCATCGAAGAGCTGATCAATTAATAT 562  
Qy 154 AepGlyLySerValIhIleGlyAepArgIu-----CysGlyPheIleAenAepGlu 169  
Db 563 GATGGCAAAAGTCCACGGTGAAGAAAGATGATCCCTGATCTGTGCTGATCAGCATGCT 622  
Qy 170 ILePheValIgluLeuValAenIaleuGlyInglInTyrAenAepAepAepAepAepAep 187  
Db 623 GTGTTTCTGAGCTGTGTGATGCTCCCTCAACAGATCTGTGATGAGAGAGAGAGAGAG 682  
Qy 188 -----AepAepGly-----AepAepProGluGlu-----ArgGlu 197  
Db 683 AACCAACCCCTCGATGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 742  
Qy 198 GluLySerGluLyAepLeuGluAepIhIleArgAepAepLyGluSerArgProAepLy 217  
Db 743 AGAAGCCCACTGCTATGAAAGCAACAA-----AAGAGTTCCAG-----AAGAG 790  
Qy 218 PheProSerAepLyIlePheGluIaIleSerSerMetPheProAepLyGlyThrIa 237  
Db 791 TTTCCAAATGACATGATCTTCAGGCGCATGTGCTCATGTTTCCGAAAGAGGTGCTCT 850  
Qy 238 GluGluLeuLySerGluLyTyrLyGluGluLeuThrgluGluIngluProGlyValAe 257  
Db 851 GACCAATGAAGAGAGGTATCGAGCTGACAGAGATGTCAGACCCCAATGCACTTCC 910  
Qy 258 ProGluCyThrProAenIleAepGlyProAenIaleuSerValGluArgGluInSer 277  
Db 911 CCTCAGTGCACACCCCAATGATGATGCCCAAGCCCAAGTCAAGTCAAGGAGAGAGCT 970  
Qy 278 LeuHleSerPheHleThrLeuPheCyAargArgCyPheLyTyrAepCyPheLeuHle 297  
Db 971 CTGCACTCTTCCACACCTTTTCTGCGCGCTGTTTAAATNGATGCTGCTCTTCAAC 1030  
Qy 298 ProPheHleAalThrProAenThrTyrLyAargLyAenThrgluThrIaleuAepAen 317  
Db 1031 CCTTCCAGCCCAACCCCAATGATATATAGCGCAAGAACCAAGAAATCAAGATGAGCA 1090  
Qy 318 LySerProGlyProGluCyThrGluHleuGluGluValAlyGluPheAalAalA 337  
Db 1091 GAACCTGTGACAGACTGCTCTTCTTGTGGAAGAGCAAGAGATGAGCC----- 1144  
Qy 338 LeuThralaGluArgIleLyThrProProLyAargProGlyIaArgArgArgIlyArg 357  
Db 1145 -----ATCTGACACAACCTCGTGCAGTGCCTCGGGCGCGCGCGAGAG 1192  
Qy 358 LeuProAenAenSerSerArgProSerThrProThrIleAenValLeu--GluSerLy 376  
Db 1193 CACCAAGGTGACAGTGTCTCTGCTCAATGATGATGATGATGATGATGATGATGATG 1252  
Qy 377 AepThrAepSerAepArgGluAlyGlyThrgluThrgluGluAenAenAepLyGlu 396  
Db 1253 GAAGGAGACAGTGAATGAGACCTGCG----- 1279

Qy 397 GluGluGluLyLyAepGluThrSerSerSerSerGluIaAenSerArgCyGluInThr 416  
Db 1280 -----AATGACTGGGCTTCAGATTTTCAGAGGCTAACTTCCTGTGACAGC 1327  
Qy 417 ProIleuMetLyAepProAen-----IleGlu---ProPro 427  
Db 1328 CCCAGCAACAGAAAGCCAGCCAGCCCAAGCTCAGCTGTGTGTGTGGAAGCCCTTCA 1387  
Qy 428 GluAenValIgluTPSerGlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 447  
Db 1388 GAGCGGTGGAATGAGACCGAGCGGAAGATCTCTTCCAGAGCTTCCAGGACCTAT 1447  
Qy 448 TyrAepAenPheCyAalIaleAlyGluLeuIleGlyThrLyThrCyAargGluValTyr 467  
Db 1448 TTCAAACAATCTGCTCAATAGCCAGGCTTGTGGGCAAGAAATCAAGATCAAGAGCTTT 1507  
Qy 468 GluPheArgVallyGluSerSerIleIleAalProAalProAalagluAepValAepThr 487  
Db 1508 CAGTTTCAGTCAAGAA-----TCACTTATCTTAAAGCTGCCACAGATGACTATGAC 1564  
Qy 488 ProProArgLyLyLyLyAargLyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 507  
Db 1565 CTTGCACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1624  
Qy 508 LeuLyLyAepArgLySerSerAenHleValIyThrAenTyrGluProCyAepAepProArg 527  
Db 1625 CTGAAGAAAGATMACAATTTTCAACAGGTGATTAATCAACAACTTGTGACACCCAGAC 1684  
Qy 528 GluProCyAepSerSerCyProCyValIleAlyAlyAlyAlyAlyAlyAlyAlyAly 547  
Db 1685 GTTCGGGTGACACACATGCTCCCTGCATGATGACCCAGAACTTTTGTGAAAGTTCTGC 1744  
Qy 1745 CAGTGCAGCCAGACCTGCAGAAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1804  
Db 568 ThrLySerGluProCyGlyTyrLeuAlyAlyAlyGluCyAepProAepLeuCyLeuThr 587  
Qy 1805 AACCAAGCAATGCTCATCTACTTGGCAGTTCGTAGGTGTACCTGACTGTGTGCTCAC 1864  
Qy 588 CyGluValIaIaAepHleSTPASPSeLyAenValSerCyLyAenAenCySerIleGlu 607  
Db 1865 TGTGGGCTTCAGAGCACTGGGACTGTAAGGTGTGCTTGCAGAAATGACAGATCCAG 1924  
Qy 608 ArgLySerLyLyLyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 627  
Db 1925 CGTGGCTCAAAAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1984  
Qy 628 ILeuLyAepProValIgluLyAenGluPheIleSerGluTyrCyGlyGluIleIleSer 647  
Db 1985 ATCAAGAGATCTGCGAAGAAATGAATTCATTTCTAAMATTTGTGGAGCTCATCTCT 2044  
Qy 648 GluAepGluIaAepAargGlyLyValIyThrAepLyTyrMetCySerPheLeuPhe 667  
Db 2045 CAGATAGAGCTGATGCGAGAGAGCTATATATATATATATATATATATATATATATAT 2104  
Qy 2105 AACCTCAACATGATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2164  
Db 688 AenHleSerValAepProAenCyTyrAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 707  
Qy 2165 AACCATTCAGTGAACCCCACTGTTATGCCAAATGTGTATGAGTGAATGAGATCACCG 2224  
Db 708 IleGlyIlePheAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 727  
Qy 2225 ATTGAGATCTTTCAGAGAGAGCAATTCAGGCTGCGAAGGCTCTTCTTGAATTAAGG 2284  
Db 728 TyrSerGluIaAepAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAlyAly 745  
Qy 2285 TACAGCAAGCTGATGCTTCAAGTATGTGGATGTCAGAGGAAACGAGAGCTC 2338

RESULT 6  
AKI40694  
LOCUS DEFINITION  
AKI40694 4194 bp mRNA linear HTC 21-SEP-2005  
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:B93092M06 product:enhancer of zeste  
homolog 1 (Drosophila), full insert sequence.

ACCESSION  
AKI40694  
VERSION  
AKI40694.1 GI:74180279  
KEYWORDS  
HTC, CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
PUBMED  
10349636  
REFERENCE  
AUTHORS  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED  
11042159  
REFERENCE  
AUTHORS  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Watanabe, K.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multiplexed sequencer  
JOURNAL  
Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED  
11076861  
REFERENCE  
AUTHORS  
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Aizawa, K., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
Arakawa, T., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nakaide, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarrelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fleischer, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P.,  
Rising, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Sessa, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toto, O. K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, Y.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kottaki, S.  
and Hayashizaki, Y.  
CONSTRM  
RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
TITLE  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
Nature 409 (6821), 685-690 (2001)  
PUBMED  
11217851  
REFERENCE  
AUTHORS  
5 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nakaide, I., Osoo, N., Saito, R., Suzuki, H., Yamana, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarrelli, R., Hill, D. P., Bult, C.,  
Hume, D. A., Quackenbush, J., Schriml, L. M., Knapin, A., Matsuda, H.,  
Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,  
Chochalski, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,  
Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S.,

Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, B. D., Kanai, A.,  
Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,  
Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
Maltais, L., Marchionni, L., McKensie, L., Miki, H., Nagashima, T.,  
Nunata, K., Okido, T., Pavani, M. J., Petter, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramchandran, S.,  
Ravelin, A., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Sempke, C. A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
Verardo, R., Wagner, L., Wanstedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yangisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
CONSTRM  
FANTOM Consortium  
TITLE  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL  
Nature 420 (6915), 563-573 (2002)  
PUBMED  
12466851  
REFERENCE  
AUTHORS  
6 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C.,  
Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R.,  
Shimokawa, K., Bajic, V. B., Bremner, S. B., Batalov, S., Forrest, A. R.,  
Zavalan, M., Davis, M. J., Walimig, L. G., Aldred, A., Allen, J. E.,  
Amesal, Implicato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L.,  
Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M.,  
Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R.,  
Crome, M. L., Dalla, E., Dalrymple, B. P., de Bono, P., Della Gatta, G.,  
di Bernardo, D., Down, T., Engstrom, P., Fegolimi, M., Faulkner, G.,  
Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M.,  
Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E.,  
Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N.,  
Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T.,  
Jakt, M., Knapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H.,  
Kitano, H., Kollas, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K.,  
Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J.,  
Liu, S., McWilliam, S., Madan, Babu, M., Madger, M., Marchionni, L.,  
Matsuda, H., Matuzawa, S., Miki, H., Mignone, F., Miyake, S.,  
Morris, K., Mortuza, Tabar, S., Mulder, N., Nakano, N., Nakachi, H.,  
Ng, P. P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O.,  
Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavoni, G.,  
Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z.,  
Ringwald, M., Roet, B., Ruan, Y., Salzberg, S. L., Sandelin, A.,  
Schneider, C., Schonbach, C., Sekiguchi, K., Sempke, C. A., Sessa, L.,  
Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D.,  
Sincalir, B., Sperling, S., Stupka, B., Sugita, K., Sultana, R.,  
Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S.,  
Tegener, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R.,  
Wei, C. L., Yagi, K., Yamashita, H., Zabarovsky, E., Zhu, S., Zimmer, A.,  
Hilde, M., Bult, C., Grimond, S. M., Teasdale, R. D., Liu, E. T.,  
Brusic, V., Quackenbush, J., Wahlestedt, C., Wickett, J. S., Hume, D. A.,  
Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, Katayama, M.,  
Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M.,  
Kato, T., Kawai, H., Kawagashira, N., Kawashima, T., Kojima, M.,  
Kondo, S., Kono, H., Nakano, K., Niimura, N., Nishio, T., Okada, M.,  
Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K.,  
Watanabe, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.  
CONSTRM  
FANTOM Consortium  
TITLE  
The transcriptional landscape of the mammalian genome  
JOURNAL  
Science 309 (5740), 1559-1563 (2005)  
PUBMED  
16141072  
REFERENCE  
AUTHORS  
7 Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakashima, M.,  
Nakamura, M., Nishida, H., Yagi, C. C., Suzuki, M., Kawai, J., Suzuki, H.,  
Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T.,  
Pang, K. C., Hallinan, J., Matlick, J., Hume, D. A., Lipovich, L.,  
Batalov, S., Engstrom, P. G., Mizuno, Y., Faghili, M. A., Sandelin, A.,  
Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and



QY	397	GIUGLUGIULYSLYASPGIUTHZSEGESSESSERGIUALAASNSERAGYSGIINTH	416
Db	1307	-----AARGACTGGGCGCTCCAGTCTTCAGAGGCTAACTCGCTGTACAGCC	135
QY	417	ProIlelymElyPProAsn-----IleGlu--ProPro	427
Db	1355	CCCAAGCAAGAAAGCAAGGCTCAGCCCGACTCGTGTGTGGAAAGCCCCCTCA	1414
QY	428	GIuSnVnIaGIuTmPserGIyAlaGIuIaIseMetPheargValIleuIegIyThr	447
Db	1415	GAGCCGGGAGATGAGCCGAGCCGAGAAATCTTTCOCAGTCTTTCACGCGACTAT	1474
QY	448	TyrASpAsnPhcYsaIaIleAlaArgLeuIleGIyThrIlySerThCYaArgGIuValYyr	467
Db	1475	TTCAACAACCTCTGCTCAATAGCCAGGCTTCTGGGAGCAAGACATGCAAGCAGGCTTT	153
QY	468	GIuPheargValyVgIuSerSerIleIleAlaProAlaProAlaGIuSphValAspThr	487
Db	1535	CAGTTTGACGTCAAGAA---TCACCTTTCCTAAAGCTGCCAAGCAAGAGGCTGTGAC	1591
QY	488	ProProArgIySLySLyArgIySLyIaArgLeuTPAlaAlaHicCYaArgIySLIegIn	507
Db	1592	CCTGCACGAAAGAAAGAAAGAAACAGAGGTTGTGGCGGCACATCGCAGAAATTGAG	1655
QY	508	LeuIySLyASPGIYSerSerAsnHiSValTYraenTYGIInProCYaSpHiAProArg	527
Db	1652	CTGAAGAAGATTAACAATTCTACACAATGTATATCTCAACCCGTGACCAACCAAG	1711
QY	528	GIInProCYaSPserSerCYaProCYaValIleAlaGInaSnPhCYaGIuIySphcYS	547
Db	1712	CGTCCGTGTGACGACATGCCCCCTGCATCATGACCCAGAACTTTGTGAAAAGTTGTGC	1771
QY	548	GIInCYSerSerGIuCYaGIuInaSnArgPheProGIYCYaArgCYaSLyAlaGIInCYaSn	567
Db	1772	CAGTGCAGCCAGACTGCAGAGATGCGTTTCGTGTGTGCGTGTAAAGACTCAGTCAAT	1833
QY	568	ThrlYsgInCYaSProCYaTYrIleuAlaValArgGIuCYaSPProAspLeuCYaLeuthr	587
Db	1832	ACCAAGCATGTCCATGTCTACTGTGCAGTTCGTGAGTGTGACCTCGACTGTGGCTCAC	1891
QY	588	CysGIyAlaAlaAspHiSTripAspSerIyAsnValIserCYySLyASnCYSerIleGIIn	607
Db	1892	TGTGGGCGCTCAGAGCAGCTGGGACTGTAAAGGTGTCTTGTGCAAAATCGCAGACTCAG	1951
QY	608	ArgGIySerIySLySHIleuIleuIeulAProSerAspValAlaGIYTrpGIyIlePhe	627
Db	1952	CGTGGCGCTCAAAAACACCTGCTGTGGCGCCCTTCGATGTGGCGGATGGGCGACCTTC	2011
QY	628	IleIyAspProValGIInIyAsnGIuIheIleSerGIuTYrCYaGIyGIuIleIleSer	647
Db	2012	ATCAAGAGAGTCTGTGCAGAAAGATAAATTCATTCTGAAATATGTGGTGAAGCTCAATCT	2071
QY	648	GIInaSPGIuAlaAspArgArgIySLyValTYraSLyIyTYrMetCYaSerPheIuPhe	667
Db	2072	CAGATGAGCTGACCTGTGAGGAGAGGCTATATATATATCATCTCCAGCTTCTCTTC	2133
QY	668	AsnLeuAsnAsnAspPheValaIAspAlaThraArgIySGIyAsnIySIleArgPheAla	687
Db	2132	AACTCAACAATGATTTGTGTGTGTATCTACCCCGAAAGGAAACAAATTCGCTTTC	2191
QY	688	AsnHiAspValAspProAsnCYaTYrAlaIySLyValIseMetValIAsnGIYAspHiArg	707
Db	2192	AACCATTCAGTGAACCCCACTGTTATGCCAAGAGGTGTTATGTGTAAATGAGATCACCG	2251
QY	708	IleGIyIlePheAlaIySLyArgAlaIleGIInThrGIyGIuIeUAspPheAspTYrArg	727
Db	2252	ATTGGAGATCTTTGCCAAGAGCAATTCAGGCTGGCGAAGAGCTCTTCTTGAATTAATAG	2311
QY	728	TyrSerGIuAlaAspAlaLeuIySLyTYrAlaGIyIleGIuArgIuIseMetGIuIle	745
Db	2312	TACGCCAAGCTGATGCTCTCAAGTATGTGTGGGACTCAGAGGAGAAACGAGCTTC	2365

RESULT 7	
CR857825	
LOCUS	CR857825
DEFINITION	Pongo pygmaeus mRNA; cDNA DKFZp469N1815 (from clone DKFZp469N1815).
ACCESSION	CR857825
VERSION	CR857825.1 GI:55726634
KEYWORDS	HTC.
SOURCE	Pongo pygmaeus (orangutan)
ORGANISM	Pongo pygmaeus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominae; Pongo. 1 (bases 1 to 4643) Koehler,K., Beyer,A., Mewes,H.W., Weill,B., Amid,C., Osanger,A., Fodor,G., Han,M. and Wiemann,S. The German CDNA Consortium Direct Submission Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuberberg, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center, Heidelberg, Germany
REFERENCE	
AUTHORS	
CONSRPT	
TITLE	
JOURNAL	
COMMENT	

## FEATURES

Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZP for ordering:  
<http://www.rzpd.de/cgi-bin/products/ci.cgi?cloneID=DKFZp469N1815>  
Further information about the clone and the sequencing project is  
available at <http://mips.gsf.de/projects/cdna/>.  
Location/Qualifiers  
1..4643

gene  
CDS

gene

CDS

```

/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp469N1815"
/tissue_type="kidney"
/clone_1db="469 (synonym: pkid1). Vector pSport1_Sfi, host
DH10B; sites SfiIa & SfiIb"
/dev_stage="adult"
/note="enhancer of zeste homolog 1 (Homo sapiens)"

1..4643
/gene="DKFZp469N1815"
119..2362
/gene="DKFZp469N1815"
/codon_start=1
/product="hypotheetical protein"
/protein_id="CAH90081.1"
/db_xref="GI:55726635"
/translation="MEIIPPTSKCITTYWRKYSEYRLRLQRLKLOANGKALYVA
NPAVKQKTOILNEBWKRLAPVQSVKPCGHPFLKCTIESIPGPAQOHMLRSL
NTVALPIVMSPLQONFVDETVLCNIPMGDEYKEDFTIELINNGYGVKVAH
BEMIIPEGSLIDAVLEIYDALINOYSDDEEBGNDTSDGKDSKEDLPTVRKRAKH
AIGNKTSKROKRPNDMIPSAIISMPEKNDPDMKRYELTRKMSPLPQCTPN
IDEPNAKSVOREOSLHSFTILFCRCCKTDCIPLPHRATNYTRYKKMLKIEBPEGG
TDFCLLEGAKYAMLNPRSKCSGRRRRRHIVASCSNASAAVAETIEBGSDDROT
GNWMASSSEANSRCQPTPKOKAPPAPOCVVEAPBEPVMTGAEBSLPRVFGYTF
NNCSIALRLGKTGCKFOFQAVESILILPDEILANPQOKKKRKLRLAAHCRKIQ
LKNDNSTOYVYVQCDHPRCDSDTSCIMTQNFCEKPCQCPDDCONRFGCGCKTQ
CNKQCCGYLAVRECDPDLCTLGTGASBHPCKVSCKNCSLORGILKHLILAPSDVAG
WGPIFKESVQAKNEFISYCEGLISQDADRGVYDYKMSFLPNLINDPVNADYVGG
NKLRFAPHSVNPNCYAKVAVNWGDRHIGTFAKAIQAGEELFLDYRTSQADLKYVGI
ERETDYL"
```

## ORIGIN

Alignment Scores:	
Pred. No.:	1,93e-229
Score:	2592.50
Percent Similarity:	76.5%
Best Local Similarity:	64.4%
Query Match:	64.0%
DB:	6
	Gaps:
	10
	Length:
	464
	Matches:
	488
	Conservative:
	92
	Mismatches:
	125
	Indels:
	53
	Gaps:
	10





QY 708 IIcglYlIephAlaIySArGAlaIleGlnThrGlyGluGluLeuPheAspTyrArg 727  
DB 2243 ATTGGATCTTTGCCAGAGGCAATTCAAGCTGGAGAGAGCTCTTCTTATTTACAG 2302  
QY 728 TYSerGlnAlaAspAlaLeuLysTyrValGlyIleGluArgGluMetGluIle 745  
DB 2303 TACAGCAGATGATGCTCTCAAGTACGTGGGATCGAGAGAGACCGACGCTC 2356  
RESULT 8  
AY402398  
LOCUS AY402398 2215 bp DNA linear GSS 15-DEC-2003  
DEFINITION Mus musculus EZH1 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY402398  
VERSION AY402398.1 GI:39758384  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
1 (bases 1 to 2215)  
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
2 (bases 1 to 2215)  
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source location/Qualifiers  
1..2215  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..>2215  
/gene="EZH1"  
/locus\_tag="HCM1210"  
ORIGIN  
Alignment Scores:  
Pred. No.: 8.11e-218 Length: 2215  
Score: 2462.50 Matches: 476  
Percent Similarity: 71.9% Conservative: 80  
Best Local Similarity: 61.6% Mismatches: 125  
Query Match: 60.8% Indels: 92  
DB: 14 Gaps: 14  
US-10-773-302-2 (1-746) x AY402398 (1-2215)  
QY 15 TTPArGlySArGValLySerGluTyrMetArgLeuArgGlnLeuLysArgPhe-ArgAr 34  
DB 43 TGGAAAGAAAGAAAGTAAATCTAGATATAGCGGCTTCACAGCTCAAGCGCTCCAGG-- 100  
QY 34 gAlaAspGluValLySerMetPheSerSerAsnArgGlnLysIleLeuGluArgThrGl 54  
DB 100 ----- 100  
QY 54 uIleLeuGlnGluTTP----- 60  
DB 101 -----CAAATATGCGGNN 150  
QY 61 ----LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGl 79

DB 151 NNN 210  
QY 79 YThrArgGluCySerValThrSerAspLeuAspPheProThrGlnValIleProLeuLys 99  
DB 211 NNN 270  
QY 99 eThrLeuAsnAlaValAlaSerValProIleMetTyrSerTyrProLeuGlnGlnAs 119  
DB 271 GTCTCTGAACACTGTGTGCACTGGTTCCTCAATGATATCTGTGTCTCTCTCCAGAGAA 330  
QY 119 nPheMetValGluAspGluThrValLeuHisAsnIleProTyrMetGluLysAspGluValLe 139  
DB 331 TTTCATGTGTGAAGATGAGAGAGGCTTTTGTGCAATATTCATATCATGGGTGAGAGGTGA 390  
QY 139 uAspGlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGln 159  
DB 391 GGAGAGAGATGAGACTTTTCATCGAAGCTGATCAATATCAATATGATGCGAAGTCCACGG 450  
QY 159 YAspArgGlu-----CyGlyPheIleAsnAspGluIlePheValGluLeuVal 175  
DB 451 TGAAGAGAGATGATCCTCGATCTGTGCTGATCAGCAATGCTGTGTCTGTGAGCTGTG 510  
QY 175 IAsnAlaLeuGlyGlnTyrAsnAspAspAspAsp-----AspAspGl 190  
DB 511 GGATGCCCTCAACACAGTACTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570  
QY 190 Y-----AspAspProGluGlu-----ArgGluGluLysGlnLysAspLe 203  
DB 571 AAAGCAAGACGACACAGACAG 630  
QY 203 uGlnAspHisIleArgAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 223  
DB 631 CGAAGGCAACAAA-----AAGAGTCCAG-----AAACGTTTCCAAAGACATGAT 678  
QY 223 ePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluLeuLysGluLys 243  
DB 679 CTTGAGCCGCAATGGCTGTCATGTTTCCGAGAAATGTGTCTCCAGACATGAGAGAGAG 738  
QY 243 eTyrLysGluLeuThrGlnGlnGlnLeuProGluValAlaLeuProProGluCyThrProAs 263  
DB 739 GTATCGAAGAGCTGACAGAGATGTGACAGCCCAATGACATCCCTCCAGTGCACACCCA 798  
QY 263 nIleAspGlyProAsnAlaLysSerValGlnArgGluLeuLeuHisSerPheHisThr 283  
DB 799 CATCGATGCCCCCAAGCGCAAGTCAAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 858  
QY 283 rLeuPheCySArGArgCySPhelLysTyrAspCySPhelLeuHisProPheHisAlaThrPr 303  
DB 859 CTTTTCGCGCGCGCGCTTTTAAATATGATGCTCTTCAACCTTCCACGCCACCC 918  
QY 303 oAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCyGlyProGl 323  
DB 919 AAATGTATATATAGCGCAAGAACAGGAATCAAGATGAGCAGAACCGTGTGCACAGA 978  
QY 323 nCySerTyrGlnHisLeuGluGluValAlaLysGluPheAlaAlaLeuThrAlaGluArgI 343  
DB 979 CTGCTTCCTTTTGTGGAAGAGCAAGAGAGTACCC-----ATGCT 1020  
QY 343 eLysThrProProLysArgProGluGlyArgArgGluLysArgLeuProAsnAspSerSe 363  
DB 1021 GCACAACTCTGGTCCAAAGTGTCTGGCGCGCGCGCGCGAGAGACCCAGAGTGTAGTGC 1080  
QY 363 rArgProSerThrProThrIleAsnValLeu-----GluSerLysAspThrAspSerAspAr 382  
DB 1081 TTCTGTGCTCAATGATGATGATGCTGTGATATGCTGTAAGTAAAGAGAGAGAGAGAGATG 1140  
QY 382 gGluAlaGlyThrGluThrGlyGlyLysAsnAspAspLysGluGluGluLysLysAs 402  
DB 1141 AGACACTGCG-----AATGA 1155  
QY 402 pGluThrSerSerSerSerGluAlaAsnSerArgCySglnThrProIleLysMetLysPr 422

Db 1156 CTGGGCGCTCCAGTTCTTCAGAGGCTAACTCTGCTGCAGACCCCGACGAAACAGAAAGC 1215  
422 oasn-----IleGlu---ProProGluAenValGluTyrPse 433  
Db 1216 CAGTCCAGCCCGACGCTGCTGTGTGTGTGAAAGCCCTCCAGAGCCGGTGGAAATGAC 1275  
433 TGLVALAGLUALaserMetPheArgValIleuIleGlyThrTyrTyrAspAenPheCysVal 453  
1276 CGAGCGCGAAGAAATCTTTTCCAGAGTCTTCCAGCGGACCTATTATTCACCACTTCTGCTC 1335  
453 aileAlaArgLeuIleGlyThrTyrThrCysArgGlnValTyrGluPheArgValValGlu 473  
1336 AATAGCCGAGCTTCTGGGACAAAGACATGAGAGCGCTCTTCAAGTTTCAGTCAGTCAGAA 1395  
473 uSerSerIleIleAlaProAlaProAlaProAlaAspThrProProArgValValVal 493  
1396 A---TCATTATCTTAAAGCTGCCAAGATGAGCTCAAGAACTTCACAGAAAGAGAA 1452  
493 sArgLysHisArgLeuTyrPalaAlaHisCysArgLysIleGlnLeuLysValYsaArgLys 513  
1453 AAGAAACACAGGTTGTGGCCGACACATGACGAAATTCAGCTGAAGAAAGATTAACA 1512  
513 rSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSe 533  
1513 TTCTACACAAATGATTAATCACTCAACCTGTGACCAACCAAGCTGCTGACGAC 1572  
533 rCysProCysValIleAlaAsnPheCysGlnLysPheCysGlnCysSerSerGluCys 553  
1573 ATGCCCTGCATCAAGACCCAGAACTTTGTGAAAGTTTGTCCAGTGCACCCGAGACTG 1632  
553 sGlnAsnArgPheProGlyCysArgCysValAlaGlnCysAsnThrLysGlnCysProCys 573  
1633 CCAAGATGCTTCTTCTGCTGTGCTGCTGAAGCTCAAGTCAATCAAGCAATGCTCATG 1692  
573 sTyrLeuAlaValAlaArgLysCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHis 593  
1693 CTACTTGCAGTTCGTGAGTGTGACCTGACTGTGCTCACTGTGAGCTGAGCTCAGAC 1752  
593 sTyrAspSerLysAsnValSerCysLysAsnCysSerIleGlnArgLysSerLysLysHis 613  
1753 CTGGAGCTGTAAAGTGCTGTCTTGCAGAAAGCTGCAGCATCCAGCTGCGCTCAAAAACA 1812  
613 sLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIlePheIleLysAspProValGly 633  
1813 CCGTCTGTGGCCCTTCGAGATGTGGCCGAGTGGGACACTTCATCAAGAGTGTGTGCA 1872  
633 nLysAsnGluPheIleSerGluTyrCysGlyGluIleIleSerGlnAsnProGluAlaAspAr 653  
1873 GAAGATGAAATTCATTTCTGAATATGTGTGAGTCACTCTCAAGATGAGGCTGATCG 1932  
653 GArgGlyLysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAspPhe 673  
1933 TCGAGGAGAGCTCTATGATTAATCATGTCCAGCTTCTTCAACCTCAACAAGATTT 1992  
673 eValValAspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnPr 693  
1993 TGTAGTGAATGCTACCCGAAAGAAACAAATTCGCTTTCGAAACCACTTCAGTGAACC 2052  
693 oAsnCystrValLysValMetMetValAsnGlyAspHisArgIleGlyIlePheAlaLys 713  
2053 CAACGTTATGCGCAAGTGTATGTATGATGAATGACATCCGCAATTTGGATTTTGGCCA 2112  
713 sArgAlaIleGlnThrGlyGluGluLeuPhePheAspTyrArgTyrSerGlnAlaAspAl 733  
2113 GAGAGCAATTCAGGCTGGCGGAAGAGCTCTTCTTATATTAATGATACAGCAAGCTGATGC 2172  
733 aLeuLysTyrValGlyIleGluArgGluMetGluIle 745  
2173 CCTCAAGTATGTGGCATCGAAGGAAACGAGACGTC 2209

RESULT 9  
AY402396  
LOCUS AY402396 2244 bp DNA linear GSS 15-DEC-2003

DEFINITION Homo sapiens EZH1 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY402396  
VERSION AY402396.1 GI:39758382  
KEYWORDS GSS.  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civejlo,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
Adams,M.D. and Cargill,M.  
1 (bases 1 to 2244)  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
2 (bases 1 to 2244)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civejlo,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
Adams,M.D. and Cargill,M.  
DIRECT SUBMISSION  
JOURNAL Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
location/qualifiers  
Source 1..2244  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
gene <1..>2244  
/gene="EZH1"  
/locus\_tag="HCM1210"

ORIGIN  
Alignment Scores:  
Pred. No.: 1,27e-217 Length: 2244  
Score: 2460.50 Matches: 471  
Percent Similarity: 72.3% Conservative: 77  
Best Local Similarity: 62.1% Mismatches: 157  
Query Match: 60.7% Indels: 53  
DB: 14 Gaps: 10

US-10-773-302-2 (1-746) x AY402396 (1-2244)

QY 15 TTPArgLysArgValLysSerGluTyrMetCArgLeuArgGlnLeuLysArgPheArg--- 33  
Db 43 TCGAAAGCAAAAGTGAATTCGAATACATGAGCTTCGACAACTTAAACGGCTTCAGGCA 102  
QY 34 ArgAlaAspGluValLysSerMetPheSerSerAsnArgLysIleLeuGluArgThr 53  
Db 103 AATATGGGNN 162  
QY 54 GluIleLeuAsnGlnGluTyrLysGlnArgIleGlnProValHisIleLeuThrSer 73  
Db 163 NNN 222  
QY 74 ValSerSerLeuArgGlyThrArgGluCysSerValThrSerAspLeuAspPheProThr 93  
Db 223 NNN 282  
QY 94 GlnValIleProLeuLysThrLeuAsnAlaValAlaSerValProIleMetTyrSerTyr 113  
Db 283 CAACATATGTTAAATGAGGTCACTGAACACAGTGCATTGGTTCATATGATATTCCTGG 342  
QY 114 SerProLeuGlnGlnAsnPheMetValGluAspGluThrValLeuHisAsnIleProTyr 133  
Db 343 TCCCTCTCCAAAGAACTTTATGTGTAGAATGAGCGGTTTGTGCAATATTCCTAC 402

```

OY 134 MetGlyAspGluValLeuAspGlnAspGlyThrPheIleGluGluLeuIleuLeuAspGlyThr
DB 403 ATGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
OY 154 AapGlyLysValIleGlyAspArgGlu-----CysGlyPheIleAsnAspGlu 169
DB 463 GATGGGAAAGTCCATGGTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 170 IlePheValGluLeuValAsnAlaLeuGlyGlnIleThrAsnAspAspAsp----- 185
DB 523 GTTTCCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
OY 186 -----AspAspAspAspGlyAspAspProGluGluArgGlu 197
DB 583 AATGACACTCGATGAAAGCAGATGACAGCAAGAAAGATCTGCAGTAACAGAAAG 642
OY 198 GlnLysGlnLysAspLeuGluAspIleAspIleAspAspLysGlyLysArgProArgLys 217
DB 643 AAGAAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 218 PheProSerAspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAla 237
DB 691 TTCCCAATAGATCATGATCTTCAGTCAATGGCTCATGTCCTGAGATGATGATGATGATGATGATG
OY 238 GlnGluLeuLysGluLysIleLysGluLeuThrGlnGlnLeuProGlyAlaLeuPro 257
DB 751 GATGACATGAAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 258 ProGluCysThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSer 277
DB 811 CCTCAGTGCACACCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 278 LeuHisSerPheHisThrLeuPheCysArgArgCysPheLysIleThrAspCysPheLeuHis 297
DB 871 CTGCACTCTTCACACACTTTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 298 ProPheIleAlaThrProAsnThrIleLysArgLysAsnThrGluThrAlaLeuAsnAsp 317
DB 931 CTTTTCATGCCACCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 318 LysProCysGlyProGlnCysIleGlnIleGlnGluGluAlaLysGluIlePheAlaIle 337
DB 991 GAACCATGTCAGACAGACTGCTTCCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 338 LeuThrAlaGluArgIleLysThrProProLysArgProGlyLysArgArgArgLysArg 357
DB 1045 -----ATCTCCACACCCCGCTCCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 358 LeuProAsnAsnSerSerArgProSerThrProThrIleAsn-----ValLeuGlnSerLys 376
DB 1093 CACCACTAGTCAAGTCTCTCTGCTCAATGCTCAAGCTCTGCTGCTGCTGCTGCTGCTGCTG
OY 377 AspThrAspSerAspArgGluIleGlyThrGluThrGlyGluLysAsnAspLysGlu 396
DB 1153 GAGGAGACAGTGAAGGAGACAGGACAGG----- 1179
OY 397 GlnGlnGluLysLysAspGluThrSerSerSerSerSerGluLysAsnSerArgCysGlnThr 416
DB 1180 -----AAGACTGGGCTCTCCAGTTCCTTCAAGAGGCTAATCTGCTGCTGCTGCTGCTG
OY 417 ProIleLysMetLysProAsnIleGluProPro----- 427
DB 1228 CCCACAAACAGAAAGCTAGTCAAGCCCACTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 428 GlnAsnValGluThrSerGlyValGluLysSerMetPheArgValLeuIleGlyThrIle 447
DB 1288 GAGCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 448 TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThrLysThrCysArgGlnValIle 467
DB 1348 TTCAACAACCTCTGTTCAATAGCCAGGCTTCTGCGGAGCCAGACCTGCAAGAGAGGCTTCTT
OY 468 GluPheArgValLysGluSerSerIleIleAlaProAlaProAlaGluAspValAspThr 487

```

---

```

DB 1408 CAGTTTGAGTGAAGAA-----TCATTTATCTCGAAGCTCCCAACAGATGATGATGATGATG
OY 488 ProProArgGlyLysLysArgLysIleValLeuThrAlaIleHisCysArgLysIleGln 507
DB 1465 CCTCAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
OY 508 LeuLysLysAspGlySerSerSerSerSerValIleThrIleIleGlnIleProCysAspIleProArg 527
DB 1525 CTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 528 GlnProCysAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 547
DB 1585 CCCCCCTGTCAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 548 GlnCysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 567
DB 1645 CAGTCAACCCAGACACTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 568 ThrLysGlnCysProCysIleLeuAlaValArgGluCysAspProAspLeuCysLeuThr 587
DB 1705 ACCAAGCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 588 CysGlyAlaAlaAspHisThrAspSerLysAsnValSerCysLysAsnCysSerIleGln 607
DB 1765 TGTGGGCTCTGAGACACTGAGACTGCAAGTGGTTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
OY 608 ArgGlySerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 627
DB 1825 CGTGACATTAAGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
OY 628 IleLysAspProValGlnLysAsnGluPheIleSerGluIleCysGlyGluIleIleSer 647
DB 1885 ATAAAGGAGTCTGTCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
OY 648 GlnAspGluAlaAspArgArgGlyLysValIleThrAspLysIleThrMetCysSerPheLeuPhe 667
DB 1945 CAGATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 668 AsnLeuAsnAsnAspPheValValAspAlaThrArgLysGlyAsnLysIleArgPheAla 687
DB 2005 AACCTCATTAATGATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG
OY 688 AsnHisSerValAsnProAsnCysIleAlaLysValMetMetValAsnGlyAspHisArg 707
DB 2065 AATCATTCAGTGAATCCCAACTGTTATCCCAAGGTCATGCTGTAAGAGAGACCATGCGG 2124
OY 708 IleGlyIlePheAlaLysArgAlaIleGlnThrGlyGluGluLeuPhePheAspThrArg 727
DB 2125 ATGGGATCTTTCAGAAAGGAGCAATTCAGACTGCGAAGAGCTCTTCTTGAATTAACAGG 2184
OY 728 TyrSerGlnAlaAspAlaLeuLysIleValGlyIleGluLysGluMetGluIle 745
DB 2185 TACGCCAAGCTGATGCTCTCAAGTACTGGGATCGAAGAGAGACCGACGATC 2238

```

RESULT 10  
 AY402397  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pan troglodytes (chimpanzee)  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Pan.  
 1 (bases 1 to 2183)  
 Todd,M.A., Tenebaum,D.M., Clevell,D.R., Lu,F., Murphy,B.,  
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smolensky,J.J.,  
 Adams,M.D. and Cargill,M.

```

TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL    Science 302 (5652), 1960-1963 (2003)
PUBMED     14671302
REFERENCE  2 (bases 1 to 2183)
AUTHORS    Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tenenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Snieksky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT    This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   location/Qualifiers
SOURCE     1. 2183
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            <1..>2183
            /gene="BZ1"
            /locus_tag="HCM1210"

ORIGIN

Alignment Scores:
Pred. No.:      9,33e-202      Length:      2183
Score:          2290.00        Matches:      431
Percent Similarity: 76.7%      Conservative: 66
Best Local Similarity: 66.5%    Mismatches: 99
Query Match:    56.5%          Indels:       52
DB:             14             Gaps:         9

US-10-773-302-2 (1-746) x AY402397 (1-2183)

Oy      105  A l a s e r v a l p r o l l e m e t t y s e r t t p s e r p r o l e u g l i n a s n p h e t v a l g l u a p 124
Db      316  G c a t t g g t t c c a t a t g a t a t t c t g g t c c c t t c c a a c a a c t t t a t g g t a c a a t 375
Oy      125  G l u t h r v a l l e u h i e a s n l e p r o t y r m e c g l y a s p g l u a l l e u a s p g l i n a s p l y t h r 144
Db      376  G a g a c g t t t t c t g c a t r i t c m t r a c a t g a g a g a t g a a g t a a g a a g a a g a t a g a c t 435
Oy      145  P h e i l e g l u l e u l e i l y s a n t y r a p g l y l y s v a l h i s g l y a s p a r g l u ----- 162
Db      436  t t t a t t a g a g a g t t a t c a a t a a c t a t g a t g g a a g t c c a t g c t g a a g a a g a g a t g a t c 495
Oy      163  -----C y s g l y p h e i l e a s n a s p g l u l e p h e v a l g l u e u v a l a s n a l e u c l y g l n 180
Db      496  C c t g g a t c c g t t c t g a t t a g t a t g t c t g t t t t t c g a g t g t g a t g c c c t g a a t c a g 555
Oy      181  T y r a s n a s p a s p a s p -----A s p a s p a s p 188
Db      556  T a c t a c a t g a g a g a g a g a g a g c a c a t g a c a c t c a g a t g a a a c a g a t a c a g c 615
Oy      189  A s p g l y a s p a s p p r o g l u a r g l u a r g l u g l u y s g l u y a s p l e u g l u a s p h i s a r g a s p 208
Db      616  A A A G A A G A T C T G C A G T A A C A G A A A G A A G A A G C A C A T G C T A T T G A A G G C A A C A A ----- 672
Oy      209  A s p l y s g l u s e r a r p r o p r o b a r g l y s p h e p r o s e r a s p l y s i l e p h e g l u a l l e s e r 228
Db      673  ---A A G A G T T C C A G -----A A A C A G T T C C C A A A T G A C A T G A T C T T C A G T C C A A T T G C C 723
Oy      229  S e r m e t p h e p r o a s p l y s g l y t h r a l a g l u g l u e u l y s g l u y s e r t y t l y s g l u e u t h r 248
Db      724  T C A A T G T T C C T G A G A A A T G T C C C A G A T G A C A T G A A G A G A G A G A G A G A G A G A G A G A G A 783
Oy      249  G l u g l i n g l u e u p r o g l y a l a l e u p r o p r o g l u y e t h r p r o a s n i l e a s p l y p r o a n 268
Db      784  G A G A T G T C A G A C C C A N N N C A C T T C C C C T C A G T G C A C C C A A C A T C G A T G G C C C C A A T 843
Oy      269  A l a l y s e r v a l g l n a r g l u g l u s e r l e u h i s e r p h e i l e t h l e u p h e c y a r g a r g 288
Db      844  G C A A G T C T G T C A C G G G A G A C A T C T G C A C T C C T T C A C A C A C T T T T T T T T G C G G C G C 903

Oy      289  C y e p h e l y s t y r a s p c y s p h e l e u h i s p r o p h e i l a l t h r p r o a s n t h r t y t l y a r g 308
Db      904  T G C T T T A A T A A C A G A C T C T T C C T T N N N C T T T T C A T G C C A C C C T A A T G A T A A A C G C 963
Oy      309  L y s a s n t h r g l u t h r a l a l e u a s p a s n l y p r o c y s g l y p r o g l i n c y t y r g l n h i l e u 328
Db      964  A A G A T A A A G A A A T C A A G A T T A A C A G A A C A A G T G C A C A G A C T C T T C T T T G C T G 1023
Oy      329  G l u g l y a l a l y s g l u p h e a l a l a l a l e u t h r a l a g l u a r g l l e l y s t h r p r o p r o l y s 348
Db      1024  G A A G A C A A A A G A G A T T G C -----A T G C T C C A C A A C C C G G C T C C 1065
Oy      349  A r g p r o g l y l y a r g a r g a r g l y a r g l e u p r o a s n s e r s e r a r p r o s e r t h r p r o 368
Db      1066  A A G T C T G T G T C T G C C C G A G A A G G C A C C A C A T A G C A T C T C T G C T C C A A T G C C 1125
Oy      369  T h r l l e a s n ---V a l l e u g l u s e r l y s a s p t h r a s p s e r a s p a r g l u a l a g l y t h r g l u 387
Db      1126  T C A G C C T C T G C T G T G C T G A G A C T A A A G A A G A G A C A G T G A C A G G A C A C A G G C ----- 1179
Oy      388  T h r g l y g l u a s n a s n a s p l y g l u g l u g l u l y s l y s a s p g l u t h r s e r s e r 407
Db      1180  -----A A T G A C T G G G C C T C A G T T C T 1200
Oy      408  S e r g l u a l a a s e r a r g c y s g l i n t h r p r o l l e l y s m e t l y s p r o a s n l l e g l u p r o p r o 427
Db      1201  T C A G A G C T A A C T C C T G C A G A C T C C C A C A A A G A A G G C T A G T C C A G C C C C A C T 1260
Oy      428  -----G l u a s n a l g l u t p s e r g l y a l a g l u a l a s e r 438
Db      1261  C A A C T G C G T A G T G A A G A C A C C T C G A G C C T G A A T G A C T G G G C T G A A G A A T C T 1320
Oy      439  M e t p h e a r g v a l l e u l e g l y t h r t y r a s n a s p h e c y a l a l e a l a r g l e u l e 458
Db      1321  C T T T T G A G C T T C C A T G A C A C C T A C T T C A A A C T T G T T C A A T A G C A G C G C T T C T G 1380
Oy      459  G l y t h r l y s t h r c y s a r g l i n v a l y r g l u p h e a r g v a l l y s g l u s e r s e r i l l e a l a 478
Db      1381  G G G A C A A A G C T G C A A G C A G C A G T T C A G T T G C A G T C A A A G A A ---T C A C T T A T C T G 1437
Oy      479  P r o l a p r o a l a g l u a s p v a l a s p t h r p r o b a r g l y s l y s l y s a r g l y s h i e a r g l e u 498
Db      1438  A A G C T C C A A A G A T G A C C T A T A G A C C C T C A C A G A A A A A A A A A A A A C A C A G A T T G 1497
Oy      499  T r p a l a a h i s c y s a r g l y l e g l i n l e u l y s l y s a s p g l y s e r s e r a s n h i e v a l t y r 518
Db      1498  T G G G C T G A C A C T G C A G A A A T T C A G C T G A A A A A G A A T A C T T T C A C A C A A G T G A C 1557
Oy      519  A s n t y r g l i n p r o c y s a s p h i s p r o a r g l i n p r o c y s a s p s e r s e r c y s p r o c y v a l l i l e 538
Db      1558  A A C T A C C A A C C C T G C A C C A C C A G C G C C C T G T G A C A C A C C C T G C C C T G A T C A T C A T G 1617
Oy      539  A l a g l u a s p h e c y s g l u y s p h e c y s g l i n c y s e r s e r g l u y s g l n a s n a r p h e p r o 558
Db      1618  A C T A G A A T T T C T G A A A G A T T T G C A G G C A A C C A C A C T G T C A A A T C G T T T C C C T 1677
Oy      559  G l y c y s a r g c y s l y s a l a g l i n c y a s n t h r l y s g l i n c y s p r o c y t y t l e u a l a v a l a r g 578
Db      1678  G G C T G C C T T A A G A C C C A G A C C A T G C A A T T C C A A G A A A G T C C T T G T A T C T G C A G T G G A 1737
Oy      579  G l u c y a s p p r o a s p l e u c y e l e u t h r c y s g l y a l a l a s p h i s t r p a s e r l y s a s n 598
Db      1738  G A A T G A C C C T G A C C T G T G T C A C C T G T G G G C C T C A G C A C T G G A C T G G A C T G C A A G G T G 1797
Oy      599  V a l s e r c y s l y s a n c y s e r i l l e g l n a r g l y s e r l y s l y s h i s l e u l e u a l a p r o 618
Db      1798  G T T T C C T A A A A C T G A G A T C C A G G T G C C T T A A A G A C A C T C T G C G C C C C 1857
Oy      619  S e r a s p v a l a l a g l y t r p g l y l e p h e i l e l y s a s p r o v a l g l n l y a s n g l u p h e i l e 638
Db      1858  T C T A T A T G C G C G G A T G G G G A C C T T C A T A A A G A G T C T G C A G A A A C A A A T T C A T T 1917

```

Qy 639 SerGluTyrCyseGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyr 658  
 Db 1918 TTTGAAATCTGTGTCAGCTCATCTTTCAGGATGAGGCTGATGACGCCGAAAGGCTTAT 1977  
 Qy 659 AspLysTyrMetCyseSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThr 678  
 Db 1978 GACAAATATCATGTCGACGCTTCTTCCTTCAACCTCAATTAAGATTTTGTATGATGTTACT 2037  
 Qy 679 ArgGlyGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCySTyralalys 698  
 Db 2038 CCGAAAGGAAACAAATTCGATTTGCAATTCATTCAGTCAATCCCAAGCTGATATGCCAA 2097  
 Qy 699 ValMetMetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThr 718  
 Db 2098 GTGGTCATGTGTAATGAGACATTCGATTTGGGATCTTTGCCAAGGGCAATTCAGCT 2157  
 Qy 719 GlyGluGluLeuPhePheAspTyr 726  
 Db 2158 GGCGAAGAGCTCTTCTTTGATTAC 2181

RESULT 11  
 AKI38942 2142 bp mRNA linear HTC 21-SEP-2005  
 LOCUS AKI38942  
 DEFINITION Mus musculus adult male aorta and vein cDNA, RIKEN full-length  
 enriched library, clone:A530023C07 product:enhancer of zeste  
 homolog 1 (Drosophila), full insert sequence.

ACCESSION AKI38942  
 VERSION AKI38942.1 GI:74223861  
 SOURCE HTC; CAP trapper.  
 KEYWORDS Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 PUBMED 10349636

REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 PUBMED 11042159

REFERENCE  
 AUTHORS Konno, H., Akiyama, J., Mishi, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Tashiro, H., Itoh, M.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 PUBMED 11076861

REFERENCE  
 AUTHORS Kawai, T., Shintagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Aizawa, T., Hara, A., Furumachi, Y., Konno, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
 Knehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,  
 Quackenbush, J., Schriml, L.M., Staudl, F., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Anon, H.,  
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
 Marchionni, L., Mashima, J., Mazarrelli, J., Mombærte, P., Nordone, P.,

CONSRMT  
 TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS

CONSRMT  
 TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS

Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schombach, C., Seta, T., Shibata, Y., Storch, K.P., Suzuki, H.,  
 Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,  
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kontsuki, S.  
 and Hayashizaki, Y.  
 RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409 (6821), 685-690 (2001)  
 11217851

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamada, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schombach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
 Chochoia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, J.J., Jarvis, E.D., Kanai, A.,  
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglocz, D.R.,  
 Maltate, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
 Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
 Verardo, R., Wagner, L., Walstedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Saito, K., Shiraki, T., Maki, K., Kawai, D., Aizawa, K.,  
 Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imocani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shintagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 PANTOM Consortium  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420 (6915), 563-573 (2002)  
 12466851

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C.,  
 Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodaira, R.,  
 Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R.,  
 Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, V., Allen, J.E.,  
 Ambesi-Impombato, A., Apweiler, R., Attalaya, R.N., Bailey, T.L.,  
 Barnal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M.,  
 Chiu, K.P., Choudhary, V., Christofides, A., Clutterbuck, D.R.,  
 Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G.,  
 di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulner, G.,  
 Fletcher, C.F., Fukushima, T., Furuno, M., Fukui, S., Gariboldi, M.,  
 Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E.,  
 Gustincich, S., Harbers, M., Hayashizaki, Y., Hensch, T.K., Hirokawa, N.,  
 Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T.,  
 Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H.,  
 Kitano, H., Kollias, G., Krishnan, S.P., Kugner, A., Kummerfeld, S.K.,  
 Kurochkin, I.V., Larsen, L.F., Lazarevic, D., Lipovich, L., Liu, J.,  
 Liuni, S., McWilliam, S., Madan Babu, M., Madra, M., Marchionni, L.,  
 Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S.,  
 Morita, K., Mottagui-Rad, S., Mulder, N., Nakano, N., Nakachi, H.,  
 Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O.,  
 Okazaki, Y., Orlandi, V., Pang, K.C., Pavan, W.J., Pavoni, G.,  
 Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z.,  
 Ringwald, M., Ros, B., Ruan, Y., Salzberg, S.L., Sandelin, A.,  
 Schneider, C., Schombach, C., Sekizuka, K., Sempile, C.A., Seno, S.,  
 Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D.,  
 Sincalari, B., Sperling, S., Stupka, E., Sugiyama, K., Sultana, R.,  
 Takenaka, Y., Taki, K., Tamojia, K., Tan, S.L., Tang, S., Taylor, M.S.,  
 Tegener, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, B., Verardo, R.,  
 Wei, C.L., Yagi, K., Yamashita, H., Zdobych, E., Zhu, S., Zimmer, A.,

Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imanura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Kono, H., Nakano, K., Nimomya, N., Nishio, T., Okada, M., Plesey, C., Shibata, K., Shizaki, T., Suzuki, S., Tegan, M., Waki, K., Wachihi, A., Okamura-Oho, Y., Suzuki, H., Kawa, I. and Hayashizaki, Y. 2002. The transcriptional landscape of the mammalian genome. *Science* 309 (5740), 1553-1563 (2005)

CONSTRM  
JOURNAL  
PUBMED  
REFERENCE

**AUTHORS**  
Kakeyama, S., Tomaru, Y., Kasuoka, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yano, C., Suzuki, M., Kawai, J., Suzuki, H., Camrind, P., Hayashizaki, Y., Wells, C., Fitch, M., Ravasi, T., Pang, K. C., Hallinan, J., Matlick, J., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Feghbi, M. A., Sandelin, A., Chalk, A. M., Motequl-Tabar, S., Liang, Z., Denhard, B. and Wankstedt, C.



QY	298	ProPheH1aAlaThrProAsnThrTyrLysAspGlyAsnThrGluThrAlaLeuAspAsn	317
Db	1034	CCTTTCACGGCACCCCAATATGTAATAGGCGAAGAACAGAAATCAAGATTGAGCA	1099
QY	318	LysProCysGlyProGlnCysTyrGlnHisLeuGluGluAlaLysGluPheAlaAla	337
Db	1094	GAACCGTGTGGCAGACGTCTTCTTTTGTGCGAAGGAGCAAGAAAGATGACCC-----	1147
QY	338	LeuThrAlaGluArgLleLysThrProProLysAspProGlyGlyValArgArgGlyArg	357
Db	1148	-----ATGCTGCACAACCTCGGTCCAAAGTCTCGGGCGCGCGCGAAG	1199
QY	358	LeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu---GluSerLys	376
Db	1196	CACCCAGAGGTGACAGTCTTCCGTCTCAATGATAGCTCTGCTATGGCTGAACATAA	1255
QY	377	AspThrAspSerAspArgGluAlaGlyThrGluThrGlyGluAsnAspLysGlu	396
Db	1256	GAAGGAGACAGTGAATAGACACTGGC-----	1282
QY	397	GluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGluThr	416
Db	1283	-----AATGACTGGGCTCTCAAGTCTTCAAGGCTAACTCTGCTGTGAGCC	1330
QY	417	ProIleLysMetLysProAsn-----LleGlu---ProPro	427
Db	1331	CCCAAGAAACAGAAAGCCAGTCCAGCCCAAGCTCAGCTCTGTTGTGTGAAGCCCCCTCA	1390
QY	428	GluAsnValGluTyrSerGlyValAlaGluAlaSerMetPheArgValLeuLleGlyThrTyr	447
Db	1391	GAGCCGGTGGATGAGCCGAGCCCAAAATCTCTTTTCCAGTCTTCCAGCGCACCTAT	1450
QY	448	TyrAspAsnPheCysAlaIleAlaArgLeuLleGlyThrLysThrCysArgGluValTyr	467
Db	1451	TTCAACAACCTTCTGCTCAATAGCCAGCTTCTGGGACAAAGACATGCAAGCAGTCTTT	1510
QY	468	GluPheArgValLysGluSerSerLleIleAlaProAlaProAlaGluAspValAspThr	487
Db	1511	CAGTTTGACGTCAAAAGAA---TCACTTATCTCAAGAGCCCAACAGATCACTCATGAC	1567
QY	488	ProProGlyLysLysArgLysHisIleArgLeuThrAlaAlaHisCysArgLysLleGln	507
Db	1568	CCTGCACAGAAAGAAAGAAAGAACACAGGTGTGTGGCGCACATGCGAAGAAATTAG	1627
QY	508	LeuLysLysAspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisAspArgX	527
Db	1628	CTGAAGAAAGATTAACAATTCTACACAAGTGTATACCTACCAACCTGTGACACCCAGAC	1687
QY	528	GlnProCysAspSerSerCysAspProCysValIleAlaGlnAsnPheCysGluLysPheCys	547
Db	1688	CGTCCGTGTGACAGACATGCCCCCTGCATCATGACCCCAAACTTTTGTGAAGAGTTCTGC	1747
QY	548	GlnCysSerSerSerGlnCysGlnAsnArgPheProGlyCysArgCysLysAlaGlnCysAsn	567
Db	1748	CAGTGCAGCCAGACCTGCCAAGATGCTTCTCTGTTCTCGCTGTAGACTCAAGTCAAT	1807
QY	568	ThrLysGlnCysProCysTyrIleuAlaValArgGluCysAspProAspLeuCysLeuThr	587
Db	1808	ACCAAGCAATGTCCATGCTACTGTGCAGTTGCTGTGAGTGTGACCCCTGACTTGTGCTCAC	1867
QY	588	CysGluAlaAlaAspHisTyrAspSerLysAsnValSerCysLysAsnCysSerLleGln	607
Db	1868	TGTGGGGCTCAGACACTGGGACTGTAAAGGTGGTGTCTTGCAAAATCTGCAGATCCAG	1927
QY	608	ArgGlySerLysLysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyTlPhe	627
Db	1928	CGTGGCTCAAAAGACACTGTGCTGTGCCCCCTTCCATGTGTGGCGAGATGGGGCACCTTC	1987
QY	628	IleLysAspProValGlnLysAsnGluPheIleSerGluTyrCysGlyGluIleIleSer	647
Db	1988	ATCAAGAGGTCTGTGCAGAAAGATTAATTCATTTCTGAATATGTGGTGAAGCTATCTCT	2047

CY	648	GlnApGgUuaAaaPArGaRGyIGLylLyvAlTyrAspLyStryrMeCySeSepheunphe	667
Db	2048	CAGATGAAGCGCTGATCGTCGAGGAAAGGCTCATGATTAAATACATGTCCAGCTTCCTC	2107
CY	668	AenLeuBnaEnaEapPheValVaIaSpAlaThr	678
Db	2108	AACTCAACATGATGTTTGTAAGTAGATCTACC	2140
RESULT 12			
LOCUS	AK153835	1238 bp mRNA linear HTC 21-SEP-2005	
DEFINITION		Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:C92016A05 product:enhancer of zeebe homolog 2 (Drosophila), full insert sequence.	
ACCESSION	AK153835		
VERSION	AK153835.1	GI:74150295	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	10349636		
REFERENCES	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Saeki, N., Carninci, P., Komno, H., Akiyama, U., Nishi, K., Kitsumata, T., Iashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimi, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Atkawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadote, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Klehl, P., Lewis, S., Matsuo, Y., Nkaido, I., Peeble, G., Wagner, L., Washio, T., Schriml, L.M., Steubli, F., Suzuki, R., Tomita, M., Maekawa, D., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Haegewa, Y., Kawaji, H., Kontseki, S. and Hayashizaki, Y.		
CONSRMT	RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
PUBMED	11217851		
REFERENCE	5		
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,		

Nikaio, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, T., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Gotoh, C., Gotoh, T., Baldelli, R., Hill, D. P., But, C., Hume, D. A., Quackenbush, J., Schiraldi, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., T. A., Chochua, C., Corbani, L. E., Cousins, S., Della, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gutinich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kanaji, H., Kawasawa, Y., Kedzierzki, R. M., King, B. Z., Kongaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Munata, K., Okido, T., Pavan, W. J., Petre, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sadelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, J., Wahlested, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, T., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

CONSRM  
TITLE  
FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)  
12466851

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

REFERENCE  
AUTHORS  
7  
Katsuyama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yag, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Fitch, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Fagiolini, M., Sadelin, A., Chalk, A. M., Mottaugni-Taber, S., Liang, Z., Lenhard, B. and Wahlested, C.

CONSRM  
TITLE  
Riken Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)  
16141073

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
8 (bases 1 to 1238)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imanura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Kono, H., Murata, M., Nakamura, M., Niinomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, M., Muramatsu, M. and Hayashizaki, Y.

TITLE  
JOURNAL  
Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
URL: http://Qualifiers

FEATURES  
source  
1..1238  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:C920016A05"  
/db\_xref="taxon:10090"  
/clone="C920016A05"  
/cell\_type="thymic cells"  
/tissue\_type="thymus"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="2 days neonate"  
66..\_51238  
/note="unlabeled protein product; enhancer of zeste homolog 2 (Drosophila) (MGD|MG1107940|GB|NM\_007971, evidence: BLASTN, 99%, match=1181)  
putative"  
/codon\_start=1  
/protein\_id="PAB32205.1"  
/db\_xref="GI:74150296"  
/translation="MGCTGKSKSEGPYCMKRVKSEYMRARQLRFRPRADDEVTKTPSS NROKILRTETLQEWKORLQPHVITVSLSLGTRECSVTSOLDPQVITPKTLN AAVSVPIMYSWSPLOQNFVDETVINIPMGDEVDQGTFTLEILIKYDGVKVD RECGFIDELFVELVNALGOYDDDDDDDDPDEREKQKDLDDDDDDKTCPPRK PADKIFKASIMPDKGTAELEKEKYELEEQQLGALPECTPINDGPAKSVAREQ SLSFHLTFRCRCKYDQCPFLAPHPATNTYKKTETALNKKPCGPCOYHLEAKER AALTAIRITKTPKPGRRGRGLPNNSSRPSTYTSVLASKOTDSBREAETETGSR"

CDS  
ORIGIN  
Alignment Scores:  
Pident. No.: 1.39e-179 Length: 1238  
Score: 2048.00 Matches: 381  
Percent Similarity: 99.2% Conservative: 7  
Best Local Similarity: 97.4% Mismatch: 3  
Query Match: 50.6% Indels: 0  
DB: 6 Gaps: 0

US-10-773-302-2 (1-746) x AK153835 (1-1238)

1 MetGlyGlnThrGlyLysSerGlyLysGlyProValCysThrArgLysArgValLys 20  
 66 ATGGGCGAGACTGGGAGAAATCTGAGAAAGGACCGGTTGTGGGAGAAAGGTGTAATA 125  
 21 SerGlyLysMetArgLysArgGlnLysLysArgPheArgArgAlaAspGlnValLysSer 40  
 126 TCAGAGTACATGAGCTGAGACAGCTCAAGAGGTTCAAGAGAGCTGATGAATTAAGACT 185  
 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuGlnGlnLysP 60  
 186 ATGTTTAGTTCGAATCGTCAGAAAATTTTGGAAGAACTGAACCTTAACCAAGAGTGG 245  
 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
 246 AAGCAGCGGAGATACAGCTTGCGACATCATGACTCTGTAGAGCTCATTTGGCGGGAGCT 305  
 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
 306 AGGAGTGTTCAGTCAACGATGACTTGGATTTTCCAGCAACAGTCAATCCCTTAAGACC 365  
 101 LeuAsnAlaValAlaSerValProIleMetLysSerThrSerProLeuGlnLysAsn 120  
 366 CTGAATGCACTGCGCTCGGTCCTTAATGACTCTTGCTCCCTTACCAAGAAATTTT 425  
 121 MetValGlnAspGlnThrValLeuHisAsnIleProLysMetGlyAspGlnValLeuAsp 140  
 426 ATGGTGGAGAGCAAACTGTTTACATTAATCTCTTAATATGGGAGTAAAGTTCTGGAT 485  
 141 GlnAspGlyThrPheIleGlnGlnLysLysAsnLysAspGlyLysValHisGlyAsp 160  
 466 CAGGATGGCACTTTCATTAAGAACTTAATAAAATTAATGAGGAAAGTGCATGGTAC 545  
 161 ArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGln 180  
 546 AGAGATGTGATTAATAATGATGAATTTTGTGGAGTTGTAATCTCTTGCTGTCAA 605  
 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
 606 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 665  
 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
 666 AAGAGTCTAGAGGATATGAGATGATTAAGAAACTTGGCCCACTCGGAAATTTCTGCT 725  
 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLys 240  
 726 GATAAAATATTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGCAACCGCAGAGACTG 785  
 241 LysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 260  
 786 AAGAGAAATATTAAGAACTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 845  
 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer 280  
 846 ACTCCAAACATGATGAGACCAATATGCAATCTGTTCAAGGAGCAAAAGCTTGCAATCA 905  
 281 PheHisThrLeuPheCysArgArgCysPheLysLysLysLysLysLysLysLysLys 300  
 906 TTTCATACGCTCTTCTGTCGACGATGTTTAAGTATGACTGCTTCTCATACCTTCAT 965  
 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
 966 GCACACACCCACACATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025  
 321 GlyProGlnCysLysGlnHisLeuGlnGlnValAlaLysGlnPheAlaAlaLeuThrAla 340  
 1026 GAGACACAGTGTTCACAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085  
 341 GlnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 360  
 1086 GAGCGTATTAAGACACACACTTAAGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1145

QY 361 AenSerSerArgProSerThrProThrIleLeuValLeuGlnLysSerLysAspThrAspSer 380  
 DB 1146 AACAGTACAGACACCCAGACACCCCACTCACTGCTGCTGAGACTCAAGAGATACAGACT 1205  
 QY 381 AsparGlnValGlyLysThrGlnThrGlyGlnGln 391  
 DB 1206 GACAG 1238  
 RESULT 13  
 AKI64192  
 LOCUS  
 DEFINITION  
 Mus musculus adult male hippocampus cDNA, RIKEN full-length  
 enriched library, clone: C630043P15 product: enhancer of zeste  
 homolog 1 (Drosophila), full insert sequence.  
 AKI64192.1 GI:74211195  
 AKI64192.1 GI:74211195  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitana, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multiplexed sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 11076861  
 4  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,  
 Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G.,  
 Quackenbush, J., Schriml, L.M., Stabili, F., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
 Baldarelli, R., Barsi, G., Blake, J., Boffelli, D., Bojunga, A.,  
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., But, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamita, M., Lee, N.H., Lyons, P.,  
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombereis, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
 Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,  
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschütter, S.,  
 and Hayashizaki, Y.  
 RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409 (6821), 685-690 (2001)  
 11217851  
 5  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

CONSRM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)  
12466851

CONSRM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

The transcripional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)  
16141072

6

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodaira, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, E., Allen, J.E., Ambesi-Impombato, A., Apweiler, R., Auraliya, R.N., Bailey, T.L., Banerjee, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crome, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furumasa, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojovic, T., Green, R.B., Guetlich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelsø, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.P., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morita, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavese, G., Pece, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Ros, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, K., Tan, S., Tang, S., Taylor, M.S., Takenaka, Y., Taki, K., Tammoja, K., Tan, S., Tang, S., Teylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yang, K., Yamashita, H., Zaslavsky, E., Zhu, S., Zimmer, A., Hilde, W., Bull, C., Grimmond, S.M., Teasdale, R.D., Liu, B.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, K., Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imanura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Niimura, Y., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Wachihi, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., and Hayashizaki, Y.

REFERENCE  
AUTHORS

7  
Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yag, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallman, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Fagiolini, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlestedt, C.

CONSRM  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Riken Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)  
16141073

8 (bases 1 to 2860)

Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imanura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Niimura, Y., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Wachihi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE  
JOURNAL

Direct Submission  
Submitted (14-APR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-tesg@gsr.riken.jp, URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome-gsc.riken.jp/  
URL: http://fanom-gsc.riken.jp/  
Location/Qualifiers

FEATURES  
SOURCE

CDS

1..2860  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:CG30043P15"  
/db\_xref="taxon:10090"  
/clone="CG30043P15"  
/sex="male"  
/issue\_type="hippocampus"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
111..1859  
/note="unlabeled protein product; enhancer of zeste homolog 1 (Orosophila) (MGD|MG1:109765 GB|U60453, evidence: BLASTN, 99%, match=1770)  
putative"  
/codon\_start=1  
/protein\_id="BAB37674.1"  
/db\_xref="GI:74211196"  
/translation="WRWDIASPETSCTIYWRKRYKSYRRLRLQRLQANAKAL  
YVAFVAFQRTQTLNLEWKKLRQVQPKPVSGHFLFKCTIESIFPGSDMLM  
RSGLTVALVIMSWSPLOQNFWEDETCVNPYMDVEBETIEELINNYDK  
VHGSEMIPIGSLVSDAVPLEVDALNQYDEEDGANDSDGQDSKEDLPYRKR  
KRHAIEGNKSSKQPPNDMIFSAIASPENGQYDMKRYRELTEMSPNALPPQC  
TPNIDGNMAVSQREGSLHFRHFCRCRCKYOCFLPHATPVPYRKRKEITIEE  
PCGTDCFLLEGAEVYMLNPKSCGGRRRRPVTSACSNASASAMATREGSD  
RDTNDWASSSENSSCQPTQKAPAPAOUCVAPSPPEVMTGAESLFEVFIG  
TYFNPFCSIALRLQTKOVFOFAPVESLILKLPTELMPAPQKKRKRRLWAHCR

Alignment Scores:  
Pred. No.: 1.55e-141  
Score: 1645.50  
Percent Similarity: 68.5%  
Best Local Similarity: 52.1%  
Query Match: 40.6%  
DB: 6  
Gaps: 15

US-10-773-302-2 (1-746) x AKI64192 (1-2860)

QY 15 TTPAAGLYSARGVALLYSESGLYTYMETATGLEUAGGLINLEUYYEARPHARG 34  
 DB 162 TGGAAAAGAAAAGTAAATCTGATATAGCGCTTCACAGCTCAAAAGCGCTCAAGCA 221  
 QY 35 AIAAAGLYVALLYSESEMERPHESERSEANATGGLINLYEILEUENGLUARGHGU 54  
 DB 222 AATATGGAAGCAAAAGCTCTGATATGCGCAATTTTGCAGAAAGTTCAAGAAAACCCAA 281  
 QY 55 ILEUAENGINGLUTRPLYSGLNARGARGLIEGIPROVALHIAILEUTHSERVAL 74  
 DB 282 ATCTTCATGAAAGAGTGAAGAAAGCTGCTGCGAGCTGTTCAAGCAATGAAGCCGG 341  
 QY 75 SERSELEUARGGLYTHARGGLUCYSERVALTHRSERAPLEU---ASPPHROTHR 93  
 DB 342 AGTGGGCAACCTTTCTGAAAAGTGTACCATAGAGACATTTTCCAGGGTTCACAGC 401  
 QY 94 GINVALILEPROLEUETHREUENALIAVALIASERVALPROILEMETYSETRP 113  
 DB 402 CAGATATATGATGCGGCTCTGAAACACTGTGACCTGCTCCATCATGATATTCCTGG 461  
 QY 114 SERPROLEUGINGLINAARPHMETVALGIUAERGLUTHRVALLEUHIENALIEPROTYR 133  
 DB 462 TCCCTCTCCAGCAGAAATTTTATGCTGGAAGATGAGCGGTTTGTGCAATTTCCATAC 521  
 QY 134 METGLYARGLIVALILEUAPRGILNAPRGILYTHREILLEUGLULEUILLERYAANTYR 153  
 DB 522 ATGCGTGAAGAGTGAAGAGAAAGATGAGACTTTCATGAGAGACTGATCATATATAT 581  
 QY 154 AARGLYLYVALIHISGLYASPARGLU-----CYAGLYPHEILEANAPRGILU 169  
 DB 582 GATGGCAAAAGTCCAGGTGAAGAAAGATGATCTCGATCTGTGCTGATCGCATGCT 641  
 QY 170 ILEPHEVALGLULEUVALAENALALEUGLYGINTYASNAAPSPAPSPAPSPAPSPAP 187  
 DB 642 GTGTTTCTGAGCTGTGATGCTCCCTCAACAGTACTGATGAGAGAGAGAGAGAGAGAG 701  
 QY 188 -----AARPARGLY-----AARPARPROGLU-----ARGLU 197  
 DB 702 AACGACCCCTCCGATGAAAGCAACAGACAGCAAAAGAGACCTGCGGTAAACAAGAAA 761  
 QY 198 GILUYSGILNLYASPLEUGLUNASPHISARGAPAPLYGILUSERARGPROPROARGLYS 217  
 DB 762 CGAAAGCCGATGCTATGGAAGGCAACAAA-----AAGAGTTCCAG-----AAACAG 809  
 QY 218 PHEPROSERAPPLYEILEPHEGLUNALIAIESERSEMERPHESERAPPLYEGLYTHRALA 237  
 DB 810 TTTCCAAATGACATGATCTTCAGCGCCATGCGCTCATGTTCTCGAAGATGATGCTCC 869  
 QY 238 GILUGLULEUUYSGILUYSTYLYSGILULEUTHRGILUGLINGLEUPROGLYVALILEUPRO 257  
 DB 870 GACGACATGAGAGAGAGATGAGAGCTGACAGAGATGTCAGACCCCAATGACATGCTCC 929  
 QY 258 PROGLUYSTHRPROAENILIASPGLYPROAENALALYSESERVALGIARGLIUNLINER 277  
 DB 930 CCTCAGTGACACCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 989  
 QY 278 LEUHIASERPHENIETHREUPHECYARGARGCYSPHELYEYASPCYSPHELEUHI 297  
 DB 990 CTGCACTCTTTTCAACACCTTTTCTGCGCGCGCTTTTAAATATGATGATGATGATGATGAT 1049  
 QY 298 PROPHENISALATHRPROASANTHYRILYARGLYASANTHYRILUTHRALALEUASPHAN 317  
 DB 1050 CCTTTCACAGCCCAACAAATGATATATAGCCAGCAAGAACAGAAATCAAGATGAGCA 1109  
 QY 318 LYSPROCYSGLYPROGLINCYETRYGINHISLEUGLUGLYVALIALYGLUPHELEALIALA 337  
 DB 1110 GAACCGTGTGGAACAGACTGCTCTTTTGTGGAAGAGCAAAAGAGAGAGAGAGAGAGAGAG 1163  
 QY 338 LEUTHRALAGLUARGILEYETHTHRPROPROLYSARGPROGLYGLYARGARGARGLYARG 357  
 DB 1164 -----ATGCTGACAAACCTCGCTCCAAAGTGTCTGGGCGCGCGCGCGAGAG 1211

QY 358 LEUPROAENANSESERARGPROSETHRPROTHRILEANVALLEU---GLUSERLYS 376  
 DB 1212 CACCCAGTGTGATAGTCTCTGCTCCATATGATGATGATGATGATGATGATGATGATGAT 1271  
 QY 377 APTHRAPSESERAPARGILUALAGLYTHRGILUTHRGILUGLYLUNASNAAPLYSGILU 396  
 DB 1272 GAAGGAGACATGANTGAGACACTGG----- 1298  
 QY 397 GILUGLILUYSLYASPRGLUTHRSERSESERSEGLUALAENSERARGCYSGILINHR 416  
 DB 1299 -----AATGACTGGGCTCCAGTTCTTCAGAGGCTTAACCTGCTGCTGACAGCC 1346  
 QY 417 PROILEYMEELYPROAEN-----ILEGLU---PROPRO 427  
 DB 1347 CCCAGAAACAGAAAAGCCAGCTCAGCCAGCTGCTGTGTGTGAAAGCCCCCTCA 1406  
 QY 428 GLUBANVALGLUTRPSERGLYALAGLUALASERMETPHEARGVALLEUILEGLYTHRTYR 447  
 DB 1407 GAGCCGCTGGAATGAGACCGGAGCCGAGAAATCTTTTCCAGTCTTCACAGGACCTAT 1466  
 QY 448 TYRASPANPHECYALALIALAARGLEUILEGLYTHRYTHRYSARGGLINVALTYR 467  
 DB 1467 TTCAACAACTTCTGCTCAATGACCGAGCTTCTGGGGAACAAGACATGCAAGCAGGCTTT 1526  
 QY 468 GLUPHEARGVALYSGILUSERSERILEILEAPROALAPROALAGLUNASVALASPHR 487  
 DB 1527 CAGTTTGCAGTCAAGAA-----TCACTTATCTTAAAGCTCCCAACAGATGAGCTATGAC 1583  
 QY 488 PROPROARGLYSLYLYARGLYSISARGLEUTHRALIALAHISCYARGLYSILEGIN 507  
 DB 1584 CCTGACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1643  
 QY 508 LEUUYSLYASPRGLYSESERSEANHISVALIYRANTYRGINPROCYASPHISPROARG 527  
 DB 1644 CTGAAGAAAGATTAACAAATTTTACACAAAGTATATATACAAACCTGAGACCAACCAAGAC 1703  
 QY 528 GINPROCYASAPSERSECYPROCYVALIILEALGINANPHECYGLINUYSPHECY 547  
 DB 1704 GCTCCGTGTGACAGACCAATGCCCCCTGACATGAGACCCAGAACTTTTGTGAAGAGTTCTGC 1763  
 QY 548 GLNCYSERSEGLUCGLINASNAARPHPROGLIYCYARGCYALYS-ALAGLNCYAS 567  
 DB 1764 CAGTGCAGCCCAAGAC-----TGTAAGACACCTGCTT 1796  
 QY 567 NTHRYSGINCYSPROCYTRYLEUVALAARGGLUCYASAPPROAPLEUCYSELUH 587  
 DB 1797 TCACCATGCTACCAACAGAGGTTGGCTTG---GAGTCCCAAGCTCTTTTCTGCTGCC 1853  
 QY 587 TCYSGLYVALIALASPHISTPRAPSERLYASNAVALSERCYSLYASANCYSESERILEGL 607  
 DB 1854 CTTAGCTGCTGAGACTTTTAAATATCCCTTGTGCTGCTGAGAGCTCTTCAAGAGAG 1913  
 QY 607 NARGGLYSELYSLYHISLEUEN 615  
 DB 1914 GAAGAGTGTGATCTCCCATGTTGTG 1938

RESULT 14  
 AK045374  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS

AK045374 3639 bp mRNA linear HTC 02-SEP-2005  
 Mus musculus adult male corpora quadrigemina cDNA, RIKEN  
 full-length enriched library, clone:B23010C14 product:enhancer of  
 zeste homolog 1 (Drosophila), full insert sequence.  
 AK045374.1 GI:26337298  
 HTC; CAP trapper.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 Carninci, P. and Hayashizaki, Y.

TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	10349636
REFERENCE	
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	11042159
REFERENCE	
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, U., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Maitahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
JOURNAL	Riken Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
PUBMED	11076861
REFERENCE	
AUTHORS	5 The PANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
PUBMED	12000000
REFERENCE	
AUTHORS	6 RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the PANTOM Consortium.
TITLE	Antisense transcription in the mammalian transcriptome
JOURNAL	Science 309, 1564-1566 (2005)
PUBMED	15800000
REFERENCE	
AUTHORS	7 The PANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).
TITLE	The Transcriptional Landscape of the Mammalian Genome
JOURNAL	Science 309, 1559-1563 (2005)
PUBMED	15800000
REFERENCE	
AUTHORS	8 (bases 1 to 3639) Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sasaki, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sobabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanabe, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
JOURNAL	Direct Submission
PUBMED	
REFERENCE	
AUTHORS	Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
TITLE	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse cDNAs.
JOURNAL	Please visit our web site for further details.
PUBMED	URL: http://genome.gsc.riken.jp/ URL: http://tanom.gsc.riken.jp/ Location/Qualifiers
FEATURES	

source	1. 3639 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB:B230110C14" /db_xref="taxon:10090" /clone="B230110C14" /sex="male" /tissue_type="corpora quadrigemina" /clone_id="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 109. 1848 /note="unlabeled protein product; enhancer of zeste homolog 1 (Drosophila) (MGD MG:1097655 GB NM_007970, evidence: BLASTN, 99%, match=1755) putative"
CDS	/protein_id="BAC2334.1" /db_xref="GI:26337299" /translation="WDIASPTSKCTTYKRYKYSYMRQLKRLQAMGAKLYVA NFAKVEKTOILNEBWKQAVQVQPKVSGHPFLKCTESIFPGPSQDMLMSL NTVALVIMYSNPLOQFMVDEBYLCPVNGDVEKEDTEIELINNYGKVG EEMIPGSLVLSDAVFLVDALNOYSEEDGNDPSDKODSDKEDLPVTRKKA AIREGKSSSKKPPNDMIFSAISMPEBNGVPDMKRYRLTETMSPNPLPOCTPN IDPKNKSVYRBSGLSHFTLPCRCRYKCPHAPHAAPPVYKRYKKEIKIRBPCG TDCFLLEBGAKEYAMLNHPSKSGRRRHPVYVSAASMAATGEGSDHDT GNFMASSSSSEANSRCQPTPKQKASPAVPAQCVVABSEPEVEMGAESLRFVHGTIF NMFCSIALRLGTYCKQVQPAKESILKLPDELMNPQKKRKRRLMAHCRKIQ LKQDNNSYGVNTPCDHPDRPCDSTCPCIMTQNFCEKQCQSPDKSTLLSPSTQV VGLGVPLTFSPAP"
ORIGIN	
Alignment Scores:	2.24e-141
Pred. No.:	3639
Score:	1645.50
Percent Similarity:	63.9%
Best Local Similarity:	47.6%
Query Match:	40.6%
DB:	Gaps: 19
US-10-773-302-2 (1-746) x AK045374 (1-3639)	
QY	15 TTPAGTGYbARGVAllySserGluTYrMetArgLeuArgGlnLeuLYsARGpHeARGArg 34
DB	151 TGGAAAAGAAAGTGAATCTGACTATATATGCGCTTGACAGCTCAACCGCTCCAGGCA 210
QY	35 AAlaSpGluVallySserMetPheSerserSerAsnArgGlnLYsIleLeuGluArgThrglu 54
DB	211 AATTATGGAGCAAAAGCTCTGTATGTGGCAATTTTCCAAAGGTTCAAAAAAACCA 270
QY	55 ILeuLeuGlnGlnIuTPlyLeGlnArgArgTleGlnProValHisIleLeuThrSerVal 74
DB	271 ATCTCATATGAAGTGGAGAAACTTCGTGTCCAGCTGTTCAAGCAATGACCCCGTG 330
QY	75 SerSerLeuArgGlyThrArgGluCysSerValThrSerSpleu---AapPheProThr 93
DB	331 AGTGGGCAACCTTTTCTGAAGAAAGTACCTAGAGGCACTTTTCCAGGCTTGACAGC 390
QY	94 GlnValIleProLeuLYsThrLeuAsnAlaValAlaSerValProIleMetYrSerTrp 113
DB	391 CAGATATGTTGATGCGCGTCTGTAAACCTTGTCATGCTGCTCCATCATGATTCCTGG 450
QY	114 SerProLeuGlnGlnIuAnPheMetValGluAspGluThrValLeuHisAenIleProTYr 133
DB	451 TCCCTCTCCAGCAAGATTTCATGTGTGAAGATGACAGCGTTTGTCCATATATCCATAC 510
QY	134 MetClyAspGluValIleLeuAspGlnAapGlyThrPheIleGluGluLeuIleLYsAenTYr 153
DB	511 ATGGGTACCGAGGTGAAGAAAGATGAGCTTTCATCGAAGCGTGAATATAACTAT 570
QY	154 AspGlyLYsValHisGlyAspArgGlu-----CysGlyPheIleAsnAspGlu 169
DB	571 GATGGCAAAAGTCCAGGTGAAGAAAGATGATATCCCTGATGTGTGTCGATCAGCGATGCT 630

QY 170 IlePheValGluLeuValAsnAlaLeuGlyGlnTyrAsnAspAspAsp----- 187  
 Db 631 GGGTTCTGGAGCTGGTGGATGCTCCCAACAGTACTCTGATGAGGAGGAGGAGCGGCGAC 690  
 QY 188 -----AspAspGly-----AspAspProGluGlu-----ArgGlu 197  
 Db 691 AACGACCCCTCCGATGGAAAGCAAGACGACGACAAAGAGACCTCGCGTACAAAGAAA 750  
 QY 198 GlnLysGlnLysAspLeuGlnLysAspHisArgAspAspLysGlnSerArgProDroArgLys 217  
 Db 751 CGGAGCGCCGATGATCGAAGGCAACAA-----AAGAGTTCCAG-----AAACAG 798  
 QY 218 PheProSerAspLysIlePheGlnLysIleSerSerMetPheProAspLysGlnThrAla 237  
 Db 799 TTTCGAATGACATGATCTTCAGCGCATGCGCTCATGCTTCCTGAGATGATGCTCCCT 858  
 QY 238 GlnGluLeuLysGlnLysTyrLysGlnLeuThrGlnGlnGlnLeuProGlnValAlaLeuPro 257  
 Db 859 GACGACATGAAAGAGAGATGAGAGCTGACAGAGATGTCAGACCCCAATGACATTCCTCC 918  
 QY 258 ProGlnCysThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSer 277  
 Db 919 CCTCAGTGCACACCAACATCGATGCGCCCAACGCGCAAGTCAGTCCAGCGGAGCAGTCT 978  
 QY 278 LeuHisSerPheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHis 297  
 Db 979 CTGCACTCTTTCACACCTTTCTTTCGCGCGCTGTTTAAATATGACGCTTCCTTCAC 1038  
 QY 298 ProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsn 317  
 Db 1039 CCTTTCACGCGCACCCCAATATATATAGCGCAAGCAAGAAATCAAGATTGAGCA 1098  
 QY 318 LysProCysGlyProGlnCysTyrGlnHisLeuGlnGlnLysAlaLysGlnPheAlaAla 337  
 Db 1099 GAACCGTGGCAGACAGCTGCTTCCTTTGCTGGAAGAGCAAGAGAGAGCC----- 1152  
 QY 338 LeuThrAlaGlnArgIleLysThrProProLysArgProGlnGlyArgArgArgLysArg 357  
 Db 1153 -----ATGCTGCACACCTCGCTCGTCCAGTCTCGTCCGCGCGCGCGCAAGG 1200  
 QY 358 LeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu-----GlnSerLys 376  
 Db 1201 CACCCAGAGCTGACAGCTTCCTCCGCTCCATGATGATGATGCTGCTGATGCTGAATCAA 1260  
 QY 377 AspThrAspSerAspArgGlnAlaGlyThrGlnArgLysGlnLysAsnAspLysGln 396  
 Db 1261 GAAGGAGACAGTATGAGACACTGGC----- 1287  
 QY 397 GlnGlnGlnLysLysAspGlnLysThrSerSerSerSerGlnAlaAsnSerArgCysGlnThr 416  
 Db 1288 -----AATGACTGGCGCTCCAGTTCTTCAGAGCGCTTCCTCGCTGTCAGACC 1335  
 QY 417 ProIleLysMetLysProAsn-----IleGlu-----ProPro 427  
 Db 1336 CCCAGGAAACAGAAAGCCAGTCCAGCCAGCTGCTGTTGTGGAAACCCCTCA 1395  
 QY 428 GlnAsnValGlnTyrSerGlyAlaGlnAlaSerMetPheArgValLeuIleGlyThrTyr 447  
 Db 1396 GAGCGGTGAGATGAGCCGAGCGCAAGATCTTTCCAGAGCTTCACAGGACCTAT 1455  
 QY 448 TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThrLysThrCysArgGlnValTyr 467  
 Db 1456 TTCAACAACCTCTGCTCAATAGCCGAGTTCGGGAGCAAGACATGCAAGCAGGCTCTTT 1515  
 QY 468 GluPheArgValLysGlnSerSerIleIleAlaProAlaProAlaGlnAspValaAspThr 487  
 Db 1516 CAGTTTGACGTCAAGAA-----TCACTTATCTTAAAGCTCCCAACAGATGAGCTATGAAAC 1572  
 QY 488 ProProArgLysLysLysArgLysHisArgLeuThrAlaAlaHisCysArgLysIleGln 507  
 Db 1573 CCTGCACAGAAAGAAAGAAACACAGGTTGTGGCGCGCACACTGCAGAGAAATTCAG 1632

QY 508 LeuLysLysAspGlySerSerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArg 527  
 Db 1633 CTGAGAAAGATTAACATTTTACACAGATGATATTAATCTTCCAACTTGAGACACCCAGAC 1692  
 QY 528 GlnProCysAspSerSerCysProCysValIleAlaGlnAsnPheCysGlnLysPheCys 547  
 Db 1693 CGTCGCTGACAGCAGCATGCGCCCTGCATCATGACCCAGAACTTTGTGAAAAGTTCTGCC 1752  
 QY 548 GlnCysSerSerGlnCysGlnAsnArgPheProGlnCysArgCysLysAlaGlnCysAs 567  
 Db 1753 CAGTGCACCCCAAGC-----TGTAAAGCACCCTGCTT 1785  
 QY 567 nThrLysGlnCysProCysTyrLeuAlaValArgGlnCysAspProAspLeuCysLeuThr 587  
 Db 1786 TCACCATGCTTACCCCAAGTGTGGCTTG-----GAGTGCACAGCTTTTCTTCCTGCC 1842  
 QY 587 rCysGlnAlaAlaAspHisThrAspSerLysAsnValSerCysLysAsnCysSerIleGln 607  
 Db 1843 CCTTAGCTGCGCGAGACTTTTAAATATCCCTTGTGTCGTCGAGGATCTCAGAGGAG 1902  
 QY 607 nArgGlySerLysLysHisLeuLeuAlaProSerAspValAlaGlyTyrGlyIlePhe 627  
 Db 1903 GAAAGTGTGATCTCCCAT-----GTTGGGGGGTGTGCTCTGTT 1941  
 QY 627 eLleLysAspProValGlnLysAsnGlnPheIleSerGlnTyrCysGlnLysIleIleSe 647  
 Db 1942 CTAT-----TTTCGATTTTAAATTTTAACTTCAATTATTT-- 1978  
 QY 647 rGlnAspGlnAlaAspArgArgLysValLysAspLysTyrMetCysSerPheLeuPhe 667  
 Db 1979 -----TATGTATATAGAGTTTTCCTCGCATGATGATGCTCG 2013  
 QY 667 eAsnLeuAsnAsnAspPheValAlaAspAlaThrArgLysGlnLysIleAspGly-- 686  
 Db 2014 TATTTGGGGGATGC-CTGTTGCCAAGAGGTGAGAGAGGATATAGATCCCTGGTA 2072  
 QY 687 -----AlaAsnHisSerValAsnProAsnGlyTyrAlaLysValMet 700  
 Db 2073 CTGAGATTGACAGCTTGAGCCACATGTGGGTGGTGGAGTCAATCTCAGGTCTTC 2130  
 RESULT 15  
 BW974704 859 bp mRNA linear EST 08-AUG-2005  
 LOCUS BW974704  
 DEFINITION Bw974704 full-length enriched swine cDNA library, adult peripheral blood mononuclear cell Sus scrofa cDNA clone PBL010097E08 5', mRNA sequence.  
 ACCESSION BW974704  
 VERSION BW974704.1 GI:71959387  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 859)  
 AUTHORS Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H., Okumura,N., Hamasina,N. and Awata,T.  
 TITLE PEDS (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries  
 JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)  
 PUBMED 14681463  
 COMMENT Contact: Hirohide Uenishi  
 Animal Genome Laboratory, Genome Research Department  
 National Institute of Agrobiological Sciences  
 2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan  
 Tel: +81-29-838-8627  
 Fax: +81-29-838-8627  
 Email: huenishi@affrc.go.jp  
 EST project with full-length enriched cDNA libraries carried out in Animal Genome Research Program (Japan) by National Institute of Agrobiological Sciences and STAFF-Institute  
 Single pass sequencing of clones derived from oligo-capped cDNA library



Vector sequences were eliminated by RepeatMasker version 2002/07/13  
and crossmatch version 0.990319  
Low quality bases were trimmed based on the quality values.

FEATURES

SOURCE

Location/Qualifiers  
1..859  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone="PBL01097B08"  
/tissue\_type="peripheral blood mononuclear cell"  
/dev\_stage="adult"  
/clone\_lib="full-length enriched swine cDNA library, adult  
peripheral blood mononuclear cell"

ORIGIN

Alignment Scores:

Pred. NO.: 4.6e-137 Length: 859  
Score: 1589.00 Matches: 282  
Percent Similarity: 99.0% Conservative: 1  
Best Local Similarity: 98.6% Mismatches: 3  
Query Match: 39.2% Indels: 0  
DB: 4 Gaps: 0

US-10-773-302-2 (1-746) x BM974704 (1-859)

QY 420 MetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMet 439  
DB 2 ATGAAGCCAAACGCTGACCCGACCGAGAACGTGGAGTGGCGCCGAGCCCTCCATG 61  
QY 440 PheArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGly 459  
DB 62 TTCAGGGTCCATCGGCACGACTACGACAACTTGTGCCATTGCGAGTTGATCCGG 121  
QY 460 ThrLysThrCysArgGluValTyrGluPheArgValLysGluSerIleIleAlaPro 479  
DB 122 ACCAAACATGTCGCGCAGGTGTATGATTTCAAGTCAAAAGATCCAGTCAACGCGCT 181  
QY 480 AlaProAlaGluAspValAspThrProProArgLysLysArgLysAsnIleArgLeuTrp 499  
DB 182 GCCCCACCGAGATGTGACACTCTCCCGAAGAAAGAGAGAAACCCGTTGTGG 241  
QY 500 AlaAlaHisCysArgLysIleGluLeuLysAspGlySerSerAsnHisValTyrAsn 519  
DB 242 GCTGCACACTGCAGAAAGATACAACTGAAAGAGATGCTCTCCACACGTTTACAC 301  
QY 520 TyrGlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAla 539  
DB 302 TATCAGCCCTGACACCCGACGCGCTTGCAGACGTTGCGCCCTTGATGATGCA 361  
QY 540 GlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGly 559  
DB 362 CAAATTTTGTGAAAGTTTGTGATGCACTTCAAGTGTCAAGACCGCTTCCGGA 421  
QY 560 CysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLys 579  
DB 422 TGCCCTTGCAAGCCGACGTCAGACAGAGCAGTCCCGTGTATCTGCTGTCCAGAG 481  
QY 580 CysAspProAspLeuCysLeuThrCysGlyAlaIleAspHisLeuTrpAspSerLysAsnVal 599  
DB 482 TGTGACCTTGACCTTGCTTACTGTGTGAGCTGCTGACCATGGAGATGTAAGATGTG 541  
QY 600 SerCysLysAsnCysSerIleGlnArgLysLysLysHisLeuLeuAlaProSer 619  
DB 542 TCCTGCAAGACTGTAGCATTCAGCGGGGTCCAAAGACATTTACTGTGGCACCATCT 601  
QY 620 AspValAlaGlyTyrGlyTyrPheIleLysAspProValGlnLysAsnGluPheIleSer 639  
DB 602 GATGTGCAGGCTGGGGGATTTTCATCAAAAGATCCGTACAGAAAAATGAATTCACCTCG 661  
QY 640 GluTyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgLysValTyrAsp 659  
DB 662 GAATCTGTGGGAGATTAATTTCTCAAGTGAAGCAGACAGAAAGGGAAGGTATGAT 721

QY 660 LysTyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArg 679  
DB 722 AAATACATGTGACAGCTTCTCTTCAACTGACATGATTTTGTGGTGAATGCACCCGC 781  
QY 680 LysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
DB 782 AAGGTACAAATTCGTTTGTGCAATCACTCAGTAAATCCAACTGCTATTCGCAAAAGTG 841  
QY 700 MetMetValAsnGlyAsp 705  
DB 842 ATGATGTTATATGCGAT 859

Search completed: August 13, 2006, 17:19:32  
Job time : 7454.4 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 13, 2006, 13:21:31 ; Search time 327.311 Seconds  
(without alignments)  
6396.890 Million cell updates/sec

Title: US-10-773-302-2  
Perfect score: 4051  
Sequence: 1 MGQTKSEKBPVCMKRRVK.....RYSQADALKYVIEHEMHP 746

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 935554401 residues  
Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=xlp  
-O=/abs/ABSSWB/spool/US10773302/runat\_11082006\_140230\_16722/app\_query.fasta\_1  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=prco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs805p -USER=US10773302@cgn\_1\_1407@runat\_11082006\_140230\_16722  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_NA:  
1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1.COMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5.COMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A.COMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B.COMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7.COMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H.COMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCPUS.COMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP.COMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/R.COMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4051	100.0	2600	3	US-09-589-892B-1
2	3812.5	94.1	2476	3	US-10-104-047-1192
3	785	19.4	3148	3	US-09-699-266A-8
4	731.5	18.1	2136	3	US-09-177-249-1
5	731.5	18.1	2136	3	US-09-812-283-1
6	731.5	18.1	2136	5	US-09-071-838A-1
7	731	18.0	2315	3	US-09-061-769A-1
8	731	18.0	2799	3	US-10-176-864-16

9	722	17.8	2327	3	US-09-061-769A-3	Sequence 3, Appli
10	707	17.5	2322	3	US-09-061-769A-4	Sequence 4, Appli
11	658	16.2	1406	3	US-09-699-266A-10	Sequence 10, Appli
12	563	13.9	489	3	US-09-589-892B-5	Sequence 5, Appli
13	436	10.8	7015	3	US-09-177-249-6	Sequence 6, Appli
14	436	10.8	7015	3	US-09-812-283-6	Sequence 6, Appli
15	436	10.8	7015	5	US-09-071-838A-6	Sequence 6, Appli
16	301	7.4	260	3	US-09-640-211A-349	Sequence 349, App
17	301	7.4	260	3	US-09-640-211A-591	Sequence 591, App
18	300	7.4	11907	3	US-08-061-376-4	Sequence 4, Appli
19	298	7.4	14255	2	US-08-320-559-1	Sequence 1, Appli
20	298	7.4	14255	2	US-08-327-392-1	Sequence 1, Appli
21	298	7.4	14255	2	US-08-306-691B-55	Sequence 55, Appli
22	298	7.4	14255	3	US-08-548-860D-1	Sequence 1, Appli
23	298	7.4	14255	7	PCT-US94-04496-1	Sequence 1, Appli
24	297	7.3	2652	3	US-09-248-796A-4828	Sequence 4828, Ap
25	286.5	6.9	7890	3	US-09-949-016-2425	Sequence 2425, Ap
26	280	6.7	3217	4	US-10-094-749-270	Sequence 270, App
27	270	6.7	1452	3	US-09-876-221-1	Sequence 1, Appli
28	267.5	6.6	2732	3	US-09-589-892B-3	Sequence 3, Appli
29	261.5	6.5	1626	3	US-09-248-796A-1025	Sequence 1025, Ap
30	257.5	6.4	3102	5	US-10-131-831-9123	Sequence 9123, Ap
31	257.5	6.4	3378	3	US-09-949-016-5147	Sequence 5147, Ap
32	250	6.2	659	3	US-09-270-767-765	Sequence 765, App
33	250	6.2	659	3	US-09-270-767-16047	Sequence 16047, A
34	232	5.7	12078	3	US-09-799-451-835	Sequence 835, App
35	232	5.7	12308	3	US-09-854-133-422	Sequence 422, App
36	205	5.1	2053	3	US-09-270-767-247	Sequence 247, App
37	205	5.1	2053	3	US-09-270-767-15529	Sequence 15529, A
38	197	4.9	672	3	US-09-533-559-4844	Sequence 4844, Ap
39	197	4.9	8241	3	US-09-902-540-798	Sequence 798, App
40	187	4.6	663	3	US-09-902-540-7911	Sequence 7911, App
41	170	4.2	189560	3	US-09-949-016-17202	Sequence 17202, A
42	163.5	4.0	1755	3	US-09-620-312D-536	Sequence 536, App
43	160	3.9	441	3	US-09-640-211A-384	Sequence 384, App
44	155	3.8	292	3	US-09-640-211A-1367	Sequence 1367, App
45	155	3.8	1431	3	US-10-200-012-19	Sequence 19, Appli

#### ALIGNMENTS

RESULT 1  
US-09-589-892B-1  
Sequence 1, Application US/09589892B  
Patent No. 6689583  
GENERAL INFORMATION:  
APPLICANT: Jennewein, Thomas  
APPLICANT: Laible, Gotz  
APPLICANT: O'Carroll, Donal  
APPLICANT: Eisenhaber, Frank  
TITLE OF INVENTION: Chromatin-Regulator Genes  
FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/09/589, 892B  
CURRENT FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945, 988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(89)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (90)..(2330)

FEATURE:  
: NAME/KEY: 3'UTR  
: LOCATION: (2331) .. (2600)  
US-09-589-892B-1

Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: Gaps: 0

US-10-773-302-2 (1-746) x US-09-589-892B-1 (1-2600)

QY 1 MetGlyGlnThrGlyLeuValSerGluLeuValProValCysTrpArgValLeu 20  
DB 90 ATGGGCCGACTGGAGAGAAATCTGAGAGGAGCCAGTTTGTGGCGGAGCGTGTAAA 149  
QY 21 SerGluTrpMetArgLeuArgGlnLeuValSerPheArgValAspGluValLeuSer 40  
DB 150 TCAGAGTACATGCAGCTGAGACAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGGT 209  
QY 41 MetPheSerSerAsnArgGlnLeuValLeuGluArgTrpGluValLeuAsnGlnGluTrp 60  
DB 210 ATGTTTAGTTCCTCAATCGTCAAGAAATTTTGAAAGAAACGAAATCTTAAACGAAATGG 269  
QY 61 LysGlnArgArgGlnLeuProValHisLeuThrSerValSerSerLeuArgGlyThr 80  
DB 270 AAACGCGCAAGATACAGCTGTGACATCTGACTTCTGTAGCTCATTTGCGGGGACT 329  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValLeuLeuValThr 100  
DB 330 AGGAGGTGTTCGTGACCAAGTCACTTGATTTTCCACACAAAGTCATCCCATTAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProLeuMetTrpSerTrpSerProLeuGlnAsnPhe 120  
DB 390 CTGAATGAGTGTCTCAGTACCCATAATGTAATTTGTGCTCTCCCTACAGCAAGATTTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTrpMetGluAspGluValLeuAsp 140  
DB 450 ATGTGTGAAGATGAAGACTGTTTACATAACATCTTATATAGGAGATGAAGTTTACAT 509  
QY 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnTrpAspGlyLysValHisGlyAsp 160  
DB 510 CAGAGTGTACTTTCACTTGAAGAACTAATAAAATTAATGATGGGAAAGTACACGGGAGT 569  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGGTGAATGCCCTTGGTCAA 629  
QY 181 TyrAsnAspAspAspAspAspAspAspGluAspAspProGluGluValArgGluLeuVal 200  
DB 630 TATATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 690 AAAAGTCTGGAGGATCACCGAGATGATAAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 750 GATATAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGACACGCAAGAAACTA 809  
QY 241 LysGluLysTrpLysGluLeuThrGluGlnGlnLeuProGluValAlaLeuProGluCys 260  
DB 810 AAGGAAAAATTAAGAACTCACCGAACAAGCTCCCAAGGCGACCTTCTCTCGAATGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnIleAsnIleSer 280  
DB 870 ACCCCCAACATGATGAGCAAAATGCTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 929  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisProPheHis 300

DB 930 TTTCATACGCTTTTCTGTAGCCGATGTTTAAATATACAGCTTCTCAATCCTTTTCAT 989  
QY 301 AlaThrProAsnThrTrpLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 990 GCAAACCCCAACACTTATTAAGCGGAGAGACACAGAAACAGCTCTAGACAAACAACTGT 1049  
QY 321 GlyProGlnCysTrpTrpGlnHisIleLeuGluValAlaLeuGluPheAlaIleuThrAla 340  
DB 1050 GAGCCACAGGTTTACCGACATTTGGAGAGGCAAAAGAGTTTGTCTGTCTTCACCGCT 1109  
QY 341 GlnArgGluLeuThrProProLysArgProGluValArgArgArgGluValLeuProAsn 360  
DB 1110 GAGCGATTAAGACCCCAACCAAACTTCCAGAGGCGCCGAGAGAGACGCTTCCCAT 1169  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1170 AACAGTAGAGGCCACACACCCCAACATTAATGTCTGAATCAAAAGATACAGACAGT 1229  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAsnAspLysGluGluGluLys 400  
DB 1230 GATAGGAGACAGGGACTGAACCGGGGGAGAGAACATATATAAAGAAAGAGAGAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1290 AAAGATGAACCTTGACCTCTCTGAGCAAAATTTCTGGGTCTCAACCACTTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGluAlaLeuAsnMetPhe 440  
DB 1350 AAGCCAAATATTGAACCTCCAGAAATGTGAGAGTGGAGTGTGAGAGCTCAATGTGT 1409  
QY 441 ArgValLeuIleGlyThrTrpTrpAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1410 AGAGTCTCTATGGCAGCTTACTATGACAAATTTCTGTGCCATGTGTAGTTAATTTGGACC 1469  
QY 461 LysThrCysArgGlnValTrpGluPheArgValLeuSerSerIleIleAlaProAla 480  
DB 1470 AAACATGTAGACAGGTGTATGAGTTTAGAGTCAAACAAATCTAGCATCAATGCTCCAGCT 1529  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
DB 1530 CCGCTGAGGATGGATGATCTCTCCCAAGGAAAAAGAAAGAGAAACACCGGTGTGGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValIleAsnTrp 520  
DB 1590 GCACACTGCAGAAAGATACAGCTGAAGAAAGAGCGCTCTCTAACATGTTTACAACTAT 1649  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1650 CAACCTGTGATCATCAACGGACCTTGTGACAGTTCTGTGCTTGTGTGATGACACA 1709  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGluCys 560  
DB 1710 AATTTTGTGAAAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1769  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValArgGluCys 580  
DB 1770 CGCTGCAGAGCAGTGCACCAACCAACAGTGCCTGTCAACCTGTGTCGCGAGATGT 1829  
QY 581 AspProAspLeuCysLeuThrCysGluValAlaAspHisArgPheAspSerLysAsnValSer 600  
DB 1830 GACCTTAACCTCTGTCTTACTGTGTGAGCCGCTGACCATTTGGACAGTAAATAATGTCTC 1889  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleuLeuLeuAlaProSerAsp 620  
DB 1890 TGCAAGAACTGCAAGTATTACGCGGGGCTCCAAAAAGCATCTAATGCGGACCATCTGAC 1949  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1950 GTGCGAGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGAATTCATCTCGAA 2009  
QY 641 TrpCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValIleAspLys 660  
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACGAAGAGGAGAAAGTATGATTA 2069

QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
 DB 2070 TACATGTCAGCTTCTGTTCAACTTGAACAAATGATTTGTGTGATCAACCCGCAAG 2129  
 QY 681 G1yAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
 DB 2130 GGTAAACAATAATCGTTTGCATAATCATTCGTTAACTCGATATCCAAAGTTANG 2189  
 QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
 DB 2190 ATGGTTAACGGTGATCAACAGATAGGTATTTTGCACAGAGACCATCCAGACTGGCGAA 2249  
 QY 721 G1uLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740  
 DB 2250 GAGCTGTTTGTGATTTACAGATACAGCCAGCTGATGCCCTGAAGTATGTGGCATCGAA 2309  
 QY 741 ArgGluMetGluIlePro 746  
 DB 2310 AGAGAAATGGAAATCCCT 2327

## RESULT 2

US-10-104-047-1192  
 ; Sequence 1192, Application US/10104047  
 ; Patent No. 6943241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: NO. 6943241el full length cDNA  
 ; FILE REFERENCE: H1-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104, 047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIOR FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1192  
 ; LENGTH: 2476  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-1192

## Alignment Scores:

Pred. No.: 0 Length: 2476  
 Score: 3812.50 Matches: 706  
 Percent Similarity: 94.6% Conservative: 0  
 Best Local Similarity: 94.6% Mismatches: 1  
 Query Match: 94.1% Indels: 39  
 DB: 3 Gaps: 1

US-10-773-302-2 (1-746) x US-10-104-047-1192 (1-2476)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTyrArgLysArgValLys 20  
 DB 93 ATGGCCAGACTGGGAGAAATCTGAGAAAGGAGCCAGTTGTTGGCGGAAGGCTTAAAA 152  
 QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
 DB 153 TCAGGTACATGCGACTGAGACAGCTCAAGAGTTCAAGCGAGCTGATGAAGTAAATAGT 212  
 QY 41 MetPheSerSerAspArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTyr 60  
 DB 213 ATGTTTAACTTCACATCGTCAGAAATTTTGAAGAAACGGAATCTTAAACCAAGATGG 272  
 QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgLysThr 80  
 DB 273 AAACAGCAAGATGACAGCTGTGACATCTGACTTGTGAGCTCATTTGGCGGAGACT 332  
 QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
 DB 333 AGGAG----- 338  
 QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnGlnAsn 120

DB 338 ----- 338  
 QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
 DB 339 ---GTGGAAGATGAAACGTTTATCATTAACATTCCTTATATGGGAGATGAAGATTTTAGAT 395  
 QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyLysP 160  
 DB 336 CAGATGTGATCTTTATTTGAAAGACTATTAATAATTTATGATGGAAAGTACACGGGAT 455  
 QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
 DB 456 AGAGATGTGGGTTATATAATGATGAATTTTGTGAGTTGTGTGAATGCCCTTGCTCA 515  
 QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
 DB 516 TATAATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 575  
 QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
 DB 576 AAAGATCTGAGAGATCACCGAGATGATTAAGAAACCCCTCGAATTTCTTCT 635  
 QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
 DB 636 GATAAAATTTTGAAGCATTTTCCATATGTTTCCAGATTAAGGCGACAGCAAGAACTA 695  
 QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProGluLys 260  
 DB 696 AAGGAAATATTAAGAACTACCGACAGACGCTCCAGCGCACTTCTCTCAAGT 755  
 QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
 DB 756 ACCCCCAATGATGAGCAAAAGCTTAATCTGTGAGAGAGACAAAGCTTACATCTC 815  
 QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisAspProHis 300  
 DB 816 TTTGATACGTTTCTGTGAGGAGATGTTTAAATATGATGATGATGATGATGATGATGAT 875  
 QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 DB 876 GCAACACCCCAACACTTATAGCGGAAAGCAAGAACAGCTCTAGAACAAACCTTGT 935  
 QY 321 GlyProGlnCysTyrGlnHisLeuGlnGluValLysGluPheAlaAlaAlaLeuThrAla 340  
 DB 936 GACCAAGTGTATCCAGCACTTTGAGGAGCAAAAGGATTTGCTGCTCTCAACCGCT 995  
 QY 341 GluArgIleLysThrProProLysArgProGlyGlyArgArgArgLysArgLysProAsn 360  
 DB 996 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGAGCGCCAGAAAGAGAGCGCTTCCAT 1055  
 QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 DB 1056 AACAGTACAGAGCCAGCACCCCACTTATATGCTGAGATCAAAAGATACAGACAGT 1115  
 QY 381 AspArgGluAlaGlyThrGluThrGlyGluLysAsnAspLysGluGluGluLys 400  
 DB 1116 GATGAGGAGAGGAGGAGTGAACGGGGGAGAGAACTATTAAGAAAGAAAGAGAGAG 1175  
 QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 DB 1176 AAAGATGAATCTTGAGAGTCTCTGAGCAATTTCTCGGTTCAAACCAATTAAGATG 1235  
 QY 421 LysProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 DB 1236 AAGCCAAATATTAACCTCTGAGAAATGTGAGTGTGCTGGAAGCTCAATGTTT 1295  
 QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 DB 1296 AGAGTCTCATTTGACACTTATGACATTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1355  
 QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 DB 1356 AAACATGTATACAGGCTATGAGTTTATGAGTCAAAAGATTTACATCATGCTCCAGCT 1415

QY	481	ProAlaGluuApbValaApbThrProProAlaGlybValbValbGlybHlaaGluuTTPAla	500
Db	1416	CCCCCTGAGGATGTGGATATCTCTCCAGAGAAAAAGAGAGAAACACCGGTTGTGGCT	1475
QY	501	AlaHisCybAaGlybIleGlnLeuIlybAaPglYserSerAaHisIsvaIlyrAsnTyr	520
Db	1476	GCACATCGCAAAAGATACAGCTGTAAAAAGACGGCTCTTAAACACAGTTTACACTAT	1535
QY	521	GlnProCybAaPbHlaaPbProAlaGlnProCybAaPbSerCybProCybValIleAlaGln	540
Db	1536	CAACCGGTGATTCATCCACGGCAGCTTGTGACAGTTGTGTCCCTGTGTGTATGACAA	1595
QY	541	AaAPbCybGluIlyrPbPbCybGlnCybSerSerGluCybGlnAaAaPbPbPbPbCyb	560
Db	1596	AAATTTGTGAAAAGTTTGTCAATGTATGTTCAGAGTGTCAAAACCGCTTTCGGAGATGC	1655
QY	561	ArgCybValaIaGlnCybAaPbThrIlyrGlnCybProCybTyrIleuAlaValaIaGlnCyb	580
Db	1656	CGCTGCAAAGCACAGTGTGCACACCAAGCATGCCCGGTCTACTCGGCTGTCCAGAGTGT	1715
QY	581	AaPbProAaPbLeuCybLeuThrCybGlyAlaAlaAaPbHlaaPbAaPbSerIyAaPbValaSer	600
Db	1716	GACCTTGACCTCTGTCTACTTGTGGACCGCTGTACATTTGGGACAGTAAAAATGTGTCC	1775
QY	601	CysIyAaPbCybSerIleGlnAaGlySerIySlybHlaLeuLeuAlaProSerAaPb	620
Db	1776	TGCAGAAACCTCAGATATTCAGCGGGGCTCCAAAAAGCATTCATGTCTGGCACATCTGAC	1835
QY	621	ValIlaGlyTTPGlyIlePbHlaIlyrAaPbProValaGlnIyAaPbGlnIlePbHlaSerGln	640
Db	1836	GTGGAGGCTGGGGATTTTATTCAGAAATCTCTGGCAAAAAAATGAATTCATCTCGAA	1895
QY	641	TyrCybGlyGlnIleIleSerGlnAaPbGlnIlaAaPbArgGlyIyIyValIyTyrAaPbIyS	660
Db	1896	TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGCAGAAAGGAAAGTATGATGAATAA	1955
QY	661	TyMetCybSerPbHlaPbAaPbAaPbAaPbPbAaPbValaIaPbAlaThrArgIyS	680
Db	1956	TACATGTACACTTCTGTCTGTCAACTTGAAACATGATTTTGTGGTGAAGCAACCGCGAAG	2015
QY	681	GlyAaPbIySIIeaGrPbAlaAaPbHlaaPbSerValaPbAaPbAaPbCybTyrAlaIyValMet	700
Db	2016	GGTAAACAAATTCGTTTGGCAATATATTCGGTAAATCCAACTGCTATGCAAAAGTTATGG	2075
QY	701	MetValaAaGlyAaPbHlaaPbIleGlyIlePbHlaIyAaPbAlaIleGlnThrGlyGln	720
Db	2076	ATGGTTAACGGTGAACACAGATAGATGATTTTGTCCAAAGAGCCATCCAGACTGGCGCA	2135
QY	721	GluLeuPbPbAaPbTyrArgTyrSerGlnAlaAaPbAlaLeuIyTyrValaGlyIleGln	740
Db	2136	GAGCTGTTTTTGAATTAACAGATACAGCAGGCTGATGCCGTGAAGTATGTGGCATCGAA	2195
QY	741	ArgGluMetGlnIlePro 746	
Db	2196	AGAGAAATGAATCCCT 2213	
RESULT 3			
US-09-699-266A-8			
; Sequence 8, Application US/09699266A			
; Patent No. 6559354			
; GENERAL INFORMATION:			
; APPLICANT: Ma, Hongchang			
; APPLICANT: Morakinyo, Layo O.			
; APPLICANT: Odell, Joan T.			
; APPLICANT: Orozco Jr., Emil M.			
; APPLICANT: Rakeleki, J. Antoni			
; TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS			
; FILE REFERENCE: Bbl164 US NA			
; CURRENT APPLICATION NUMBER: US/09/699,266A			
; CURRENT FILING DATE: 2000-10-27			
; PRIOR APPLICATION NUMBER: PCT/US99/08385			
; PRIOR FILING DATE: 1999-04-16			

PRIOR APPLICATION NUMBER: 60/083,212			
PRIOR FILING DATE: 1998-04-27			
NUMBER OF SEQ ID NOS: 13			
SOFTWARE: Microsoft Office 97			
SEQ ID NO 8			
LENGTH: 3148			
TYPE: DNA			
ORGANISM: Zea mays			
US-09-699-266A-8			
Alignment Scores:			
Pred. No.:	2,17e-65	Length:	3148
Score:	785.00	Matches:	240
Percent Similarity:	40.9%	Conservative:	119
Best Local Similarity:	27.3%	Mismatches:	245
Query Match:	19.4%	Indels:	275
DB:	3	Gaps:	36
US-10-773-302-2 (1-746) x US-09-699-266A-8 (1-3148)			
Dy	47	GLNLYELEGLULARGTHGLULLEUENGLNUTRPLVSGINARGATGILEGLN	66
Db	308	CAAGGGTACTGATTCAGCATCAATTTGCTCACAAAAGCAAGATGATGCGGTATGCA	367
Dy	67	ProValHisLeu-----LeuHisValSerSerLeuArgLysThrArg	81
Db	368	CCCTGCATATGCTCTGATATATTCCGGTTGACAAAGATGCTGATCTTTCCAAACGAAA	427
Dy	82	GLUCYSERVALTHRSERAPLEUAPHE-----PROTHGINVALILE	96
Db	428	GTCCTTTCCT-TCATCTAATGTATGTTGGTGGAAATCTTGCTCCAAAGATCTAAT	486
Dy	97	-----PROLEULYTHRLEUASNAVALASERVALPROILEMCTYRSETRTPSER	114
Db	487	ATTAGACCAATTA---CTACAGAGTGGCCAAAGCTTCCACCTTATACACATGGATA	543
Dy	115	PROLEUGLGLASN--PHEMETVALGILUPSERVATHRVALLEUHSANILEPROTYR	133
Db	544	TTTTTGACGAAACCAAGATGACAGAAAGCCATTTGACTTGGTCCA-----	594
Dy	134	METGLYAPGLVALLLEUAPGLNAPGLYTHRPHLEIGLULLEULIELYEASNTYR	153
Db	594	-----	594
Dy	154	ASPGLYLYSVALHISGLYASPARGLUCYSGLYPHEILEASAPGLIILEPHEVALGLU	173
Db	595	---CGAGAGATTACATATGATCCAGTTGTGGT-----GAAGCTTAAT--	636
Dy	174	LEUVALASNALEUCLYGLINTYRANAPASPARAPASPARAPGLYASPARPRO	193
Db	637	-----TGCACTGATAGTGAAGATGAAGCCATTGAAGAT	665
Dy	194	GLUGLARGLUGLULYSGLIYASPREUGLUNAPHS-----ARGASPARPLYS	210
Db	670	GAGAGAGAAAAAGAAATTTAAACATCTGAAGATCAATTATTCGAGAGACGTTCAA	729
Dy	211	GLUSERARGPROARGLYSPHEPROSERAPLYSILEPHEGLUALILESERMET	230
Db	730	GAATGTGGC-----ATGCTGATGCTGTACGCAAGCGTAACTGCTGCACAC	774
Dy	231	PHERPAPRPLYSGLYTHRALGLULIUEULYSGIULYSTRYKLYSGIULIUEGLUGLN	250
Db	775	ATG-----GAGCGGGCTGCTGATGACATAAAGCCAGAGTATGAAATTCGACATGCTGAG	828
Dy	251	GLIUEUPROGLYALALEUPROPRO-----GLU	259
Db	829	AAAACTAAGGATTTCTTGCAAGAAAGGACTAGCATATGTCCAAAGTGAAGATTGTAC	888
Dy	260	CYSETHPROASNILEAPGLYPROASNIALYSSERVALGINARGLUGLINSERLEUHS	279
Db	889	TGTGACAAAGATTGTGATGCA-----GCATTGGAT	918
Dy	280	SERPHENHISTHRLEUPHECYARG-----ARGCYRPHELYSTRYAPRCYRPH	295

```

Db      919 TTTTGGAGAAATCTCTCTGTGCAACAGAGAAACAAGATGCTAGTGTGTTGATTCAG 978
Qy      296 LeuHis-----ProPheHisAlaThrProAsnThrTyrlvs 307
Db      979 CTACATGGGTGTTCTCAAGATTAGTATTTCCCT-----ACAGAAAA 1020
Qy      308 ArglyAsnThrGluThrAlaLeuAsnLysProCysGlyProGlnCysTyrlvs 327
Db      1021 CAACACGCTTGAGAGTGGCTGTGATGACAGTGTACCTGTGTATTCATGTCCTCAATTAAC 1080
Qy      328 LeuGluGlyAlaLysGluPheAlaAlaAlaLeuThrAlaGluArg-----Ile 343
Db      1081 -----GCATCTGAACCAAGATGCTGTGCTGTGCTGATCACAATGCTTTTGAATTT 1131
Qy      344 LysThrProPro-----LysArgProGlyGlyArgArgArg--- 355
Db      1132 GAGAGCCAACTCACTCATCAGACATGTGATGAACCAAGCCAGGTTCAATATGAGAAAAAG 1191
Qy      356 -----GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsn 371
Db      1192 AAGGCTCCAGTGGAGAGAACTAAATCTCAACAAAGTGAAGCTCTTCAACTGCACAGA 1251
Qy      372 ValLeu---GlySerLysAspThrAsp----- 379
Db      1252 GTTATCTCAAGAAAGCATGATTCGAGAAAGTACATCCATTAAGCAATTAATCTCACAACAC 1311
Qy      379 ----- 379
Db      1312 TCCCTAGTCCCTCAAAAGTTAAATTTGGCCGAAAGTGAATCAGAAAGATTACANT 1371
Qy      379 ----- 379
Db      1372 AGACGAATCGCTGAGAAATCTTATGAGTGTGAAGAAAGCAAGGAAATGGCATCA 1431
Qy      380 SerAsp-----ArgGluAlaGlyThrGluThrGly 389
Db      1432 TCTGATCTCAATTTTGTAGTGTATCTTTTGGCAGGAGCAATGAAGCTTAGCTGTGAT 1491
Qy      390 GlyGluAsnAsnAspLysGlu-----Glu 397
Db      1492 ACACGAATGGAATTAAGAAATTAATGTATCTTCACAACAGAGTTCTCCAAACAGAGA 1551
Qy      398 GluGluLysLysAspGluThr----- 404
Db      1552 AGTTCAAAAGAGAGTACACCTCAAAATTTGGAAACAGCTCAGTTTGTGAGGCTCAT 1611
Qy      405 ---SerSerSerSerGluAlaAsnSerArg----- 413
Db      1612 AATGATTTCAACAGAGAGCAAAATACCGTCAATTCAAGCAACAGATGTTACATAGTTCA 1671
Qy      414 -----CysGlnThrProIleLysMetLysProAsn 423
Db      1672 AGAAAGAAAGATTCGTCAATGAGAATTATTCAGACAGAGAGTGTGACTGAGA----- 1725
Qy      424 IleGluProProGluAsnValGluThrSerGlyAlaGluAlaSerMetCysPheArgValLeu 443
Db      1726 -----TCATGGAAAGCAATTTGACAGAGCAAGCTTCTTGTGAAGAGA 1764
Qy      444 IleGlyThrTyrlvsAspAsnPheCysAlaIleAlaArg-----LeuIleGlyThrLys 461
Db      1765 TTAGAGATTTTGAAGAGAACGTTGTTAATTTGCTGGAACTTCTGTGGAAATGAAG 1824
Qy      462 ThrCysArgGlnValTyrlvsGluPheArgValLysGluSerSerIleIleAlaProAlaPro 481
Db      1825 AGCTGCAAGATGTTTTCATATATATGATTAATTAAGAAACAAGAGCTCTGAGAGCT 1884
Qy      482 AlaGluAspValAsp-----ThrProProArgLysLys 492
Db      1885 CTAGTGTGTGATTTCTGTGTAAAGATTAATTAAGGTATGAGTGTGAGCAACAGA 1944
Qy      493 LysArgLysHisLeuArg-----LeuThrAlaAla--- 501

```

```

Db      1945 TCAAGATATTTTGAAGAGCGAGTAAAGTCCGTGTTGAAGTACACCTGAATCTGCA 2004
Qy      502 -----HisCysArgLysIleGluLeuLysLysAspGlySerSerAsnHisValTyrlvs 519
Db      2005 GGTTACAAATTTCAAAAGAGATTACCGAAGAGAGATTCAGCTTGTGCA-----CAA 2055
Qy      520 TyrlsGlnProCysAspHisLeuProArgLysProCysAspSerSerCysProCysValIleAla 539
Db      2056 TTTATCTCTTGTGCT---TGTCAATCTTACATCCGAAAGACAGTGTCCATGCTTTCATAT 2112
Qy      540 GlnAsnPheCysGlyLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGly 559
Db      2113 GGCAATGTGTGAGAAATGACTGTGGGTGTCCAAAATTTTGCAAGATCGTTTTCAGAGA 2172
Qy      560 CysArgCys---LysAlaGlnCysAsnThrLysGlnCysProCysTyrlsLeuAlaValArg 578
Db      2173 TGTCACTGTGCAAAAGCAGAGTGTGACGCCGCCAATGTCCATGTTTTCAGCTGACAGG 2232
Qy      579 GluCysAspProAspLeuCysLeuThrCys-----GlyAlaAlaAsp----- 592
Db      2233 GAATCGATCCGATGTTTTCAGAAACTGTGTGGGTGTGTGATGTGATGTCATTTGGA 2292
Qy      593 -----HisThrAspSerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySer 610
Db      2293 GTTCCAAACAGAGAGAGATTAATTAAGATGCGGAAACATGAACCTGCTTAAACA 2352
Qy      611 LysLysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrlsGlyIlePheIleLysAsp 630
Db      2353 CAACAAAGGCTTCTTCTGGAAGATCAGATGCTGTGCTGGGAGCATTTCTCAAGAT 2412
Qy      631 ProValGlnLysAsnGluPheIleSerGluTyrlsGlyGluIleIleSerGlnAspGlu 650
Db      2413 AGTGTACCAACATGAATATCTGTGTAGTACACTGGGAACTATCTCACAACAAGA 2472
Qy      651 AlaAspArgArgGlyLysValTyrlsAspLysTyrlsMetCysSerPheLeuPheAsnLeuAsn 670
Db      2473 GCAGATAGCGGTGAGAAATATATATGATCTGTGAACCTATGCTTTTCAACCTGAAC 2532
Qy      671 AsnAspPheValValAspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSer 690
Db      2533 AATGATGTGTCTTGAAGCATATCAATGTGGTGCACAGATTTGCCAACCATGCC 2592
Qy      691 ValAsnProAsnCysTyrlsAlaLysValMetValAsnGlyAspHisArgIleGlyIle 710
Db      2593 CTTGACCCGAATGTGATGCCAAGGTTATCATGTGATCTGGATCATTAAGTGGGCATTA 2652
Qy      711 PheAlaLysArgAlaIleGlnThrGlyGluGluLeuPhePheAspTyrlsArgTyrls 728
Db      2653 TTCCGCAAGAAAGAAATCTCTGCTGTGAGAGATTATTTTCAGATTACCGCTAT 2706

RESULT 4
US-09-177-249-1
; Sequence 1, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadevari, Ramim
; APPLICANT: Margosian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; EARLIER FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

```





QY 576 aValaArgGluCyAspProAspLeuCy-----LeuThrCyGlyAlaAlaAs 592  
 Db 1551 TAATCGTGAATGATCGAGATCTTTGTGAGATTCTCTCTAGCTGTGAGATGGCAC 1610  
 QY 592 pHisTrpAspSerIys---AsnValSerCysIysAsnCySerIleGlnArgIysery 611  
 Db 1611 TCTTGATGAGACACCGATGCAATTCAGACGAAGAACGCAATTCCTTCAACCAA 1670  
 QY 611 sIysHisLeuLeuLeuAlaProSerAspValAlaGlyTrpGlyIlePheIleIysAspPr 631  
 Db 1671 TTAATAAGATTCTCATTTGAAAGCTGATGATCTGATGATGGGTCATTTACATGGGACATC 1730  
 QY 631 oValGlnIysAsnGlnPheIleSerGlnIysGlyGlyIleIleSerGlnAspGlnAl 651  
 Db 1731 TCTTAAAAAGATGATATCTCGAGAAATTAATCTGAGAACTGATCATCATGATTAAC 1790  
 QY 651 aAspArgArgGlyIysValIysAspIysTrpMetCysSerPheLeuPheAsnLeuAsnAs 671  
 Db 1791 TTAATGACCTGGGAGAAATGAAAGATCGAATGGTTCTTCTTACCTTTTACCTTGAATGA 1850  
 QY 671 nAspPheValIAspAlaIleThrArgIysGlyIAsnIysIleArgPheAlaAsnHisSerVa 691  
 Db 1851 TCAGCTCGAAATCGATGCTCGCGTAAAGAAACGAGTCAATTTCTCAATCACTCAAC 1910  
 QY 691 IAsnProAsnCyTrpAlaIleValMetMetValAsnGlyAspPheIleArgIleGlyIlePhe 711  
 Db 1911 AAGACCTTAATCTCTACCGCAAGTTGATGATGATGAGAGATCAAGAGATGGCTATTT 1970  
 QY 711 eAlaIysArgAlaIleGlnThrGlyGlnIleuPhePheAspIysArgIysSerGlnAl 731  
 Db 1971 TCCGAGAGAGCAATCGAAGAAAGTGAAGAGCTTTTCTTCACTACTGCTATGACACAGA 2030  
 QY 731 aAspAla 733  
 Db 2031 ACATGCG 2037  
 RESULT 5  
 US-09-812-283-1  
 ; Sequence 1, Application US/09812283  
 ; Patent No. 6828477  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Robert L.  
 ; APPLICANT: Ohad, Nir  
 ; APPLICANT: Kiyosue, Tomohiro  
 ; APPLICANT: Yadegari, Ramih  
 ; APPLICANT: Margossian, Linda  
 ; APPLICANT: Harada, John  
 ; APPLICANT: Goldberg, Robert B.  
 ; APPLICANT: The Regents of the University of California  
 ; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
 ; FILE REFERENCE: 023070-086120US  
 ; CURRENT APPLICATION NUMBER: US/09/812,283  
 ; PRIOR FILING DATE: 2001-03-19  
 ; PRIOR APPLICATION NUMBER: 09/177,249  
 ; PRIOR FILING DATE: 1998-10-22  
 ; PRIOR APPLICATION NUMBER: US 09/071,838  
 ; NUMBER OF SEQ ID NOS: 324  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2136  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis sp.  
 ; FEATURES:  
 ; NAME/KEY: CDS  
 ; LOCATION: (43)..(2112)  
 ; OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1) cDNA  
 US-09-812-283-1  
 Alignment Scores: 1.77e-60 Length: 2136

Score: 731.50 Matches: 213  
 Percent Similarity: 42.3% Conservative: 127  
 Best Local Similarity: 26.5% Mismatches: 260  
 Query Match: 18.1% Indels: 205  
 DB: 3 Gaps: 28  
 US-10-773-302-2 (1-746) x US-09-812-283-1 (1-2136)  
 QY 6 IysIysSerGlnIysGlyProValCyTrpArgIysArgValIysSerGlnIysMetArg 25  
 Db 17 AGAAAAAGAAAGAGCGAGTGTAAATGAGAAAGAAACCATGAGACGATGGTGAAG 76  
 QY 26 Leu---Arg-GlnLeuIysArgPheArgArgAlaAspGlnValIysSerMetPheSerSe 44  
 Db 77 GTTGGCACCCGAACTAAATTCAGATTAAGACCAATGCA---AAAGAGAGATTCTGC 133  
 QY 44 IAsnArgGlnIysIleLeuGlnArgThrGlnIleLeuAsnGlnIysTrpIysGlnArgAr 64  
 Db 134 ATATCAAGAGAAATTT-----CGAGCTGAGATTA-TCCAAGTGTG---CTACTCAT 183  
 QY 64 GllleGlnProValHisIleLeuThrSerValSerSerLeuArgGly----- 79  
 Db 184 GCTTCACACCATCAATCGTTGACTTAACCAAGCCCGCTGCAAGAGATGATATGAGAGA 243  
 QY 80 -----ThrArgGlnCysSerValIleSerAspLeuAspPheProThrG1 94  
 Db 244 GACAACAAAATCACTTTGTGAGAAATGCAAAACCACTTCGATTTCACTGCTCATGT 303  
 QY 94 nValIleProLeuIysTrpIleuAsnAlaVal-----AlaSerValProIleu 110  
 Db 304 GATTATTAATTTCTTACGAAATCAAGTTATGTTCTTGATGAGAGATCAAGATTATGCTTT 363  
 QY 110 cTySerTrpSerProLeuGlnGlnAsnPheMetValGlnAspGlnThrValLeuHisAs 130  
 Db 364 GAAGAAGATGA-CCATTA-----TTCTTGATGAGATGATCAATTAATTAACAG 413  
 QY 130 nIleProIysMetGlyAspGlnValIleuAspGlnAspGlyThrPheIle----- 146  
 Db 414 TGTCAAGCTTCCAAATTTGTTGAGAGCTACCAAGATCCATTAATGAGTCTTCAACCAAG 473  
 QY 147 -GlnIleuLeuIleIysAsnIysAspGlyIysValHisIleGlyAspArgIysGlyPhe1 166  
 Db 474 TAGCAGCTGATGCGTGAAGAT---TCGTGATGTGTAAGACAAATATATATT 530  
 QY 166 eAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlnIysIysAsnAspAspAs 186  
 Db 531 GAATGATGAG-----GCACTTAAGATTAAGACAGTGAAGAAAGATCA 569  
 QY 186 pAspAspAspGlyAspAspProGlnIu---ArgGlnIysGlnIysAspLeuGlnAs 205  
 Db 570 GGAAGATGAAGAAAGATGAGAAAGAAATCAAGAAAGAAATGCGAATTTCTTGAAGA 629  
 QY 205 pHisArgAspAspIysGlnSerArgProProArgIysPheProSerAspIysIlePheG1 225  
 Db 630 T-----GTAGACC 638  
 QY 225 uAlaIleSerSerMetPheProAspIysGly----- 235  
 Db 639 ATTATATGAGACGGTTGGCAGACTATGCTGATGATGATGATGATGATGATGATGATGAT 698  
 QY 236 -----ThrAlaGlnIleuLeuIysGlnIysGlyIysGlyIysGlnLeuThrG1 249  
 Db 699 CGCCAAAGTACCTCGAAGTGAATGTTTGGACACATATGAAAGATGAATGAATGAAGCT 758  
 QY 249 uGlnGlnIleuProGlyAlaLeuProProGlnIysTrpProAsnIleAspGlyProAsnAl 269  
 Db 759 TAAGAT-----GATGAAGCTGCG 779  
 QY 269 aIysSerValGlnArgGlnIysSerLeuHisSerPheHisThrLeuPhe----- 285  
 Db 780 TGAGGCT-----TCGTATTTGACATCCAAAGCAATACTAGCTTTCCAGATTTTGC 833  
 QY 286 -----CysValArgGlyPheLeuIysTrpIleuAspGlyIysPheLeuHisVal 301

Db	834	TCATPACGCTCATTCGCGTCGTCGATCATATTCGATTCATTCATGACATGAAAGTAT	----	AspAsnIly	318
Qy	301	AThPrOaSnThrTYrLYsArGLysAsnThrGLuThrAlaLeu	----	AspAsnIly	318
Db	892	-GAGCCGAGCTGTATGATCCAGGAAAGAAATCTAGTTGTTTGAGATGAAGATGAGCA	----	AspAsnIly	950
Qy	318	sPrOCySGLyPrOGInCYsTYrGLInHLeuGLuLYaLYsGLuPheAlaAlaLe	----	AspAsnIly	338
Db	951	ACCATGCAAGTAGACATTTTACCTTCAGGTGAGAGTGCAGAA	----	AspAsnIly	996
Qy	338	uThrAlaGLuArgLLeuSThPrPrOCrOLyAsnPrOCrLYaArgArGLyArgLe	----	AspAsnIly	358
Db	997	-----GCTATCATGTG	----	AspAsnIly	1010
Qy	358	uPrOHaNsSerArGPrOsErThrPrOThrILeAnValLeuGInSerLYsAspTh	----	AspAsnIly	378
Db	1011	GGATPATATATACCT	----	AspAsnIly	1026
Qy	378	rAspSerAsnArgLualaGLyThrGLuThrGLyGLuAsnAsnAspLYsGLuGLuGL	----	AspAsnIly	398
Db	1026	-----	----	AspAsnIly	1026
Qy	398	uGLuLYsLYsAspGLuThrSerSerSerSerGLuAlaAsnSerArgCYsGLInThrPrO	----	AspAsnIly	418
Db	1027	-----AT	----	AspAsnIly	1028
Qy	418	eLYsMetLYsPrOaenILeGLuPrOCrOLuAsnValGLuTrpSerGLyAlaGLuAlaSe	----	AspAsnIly	438
Db	1029	ATCAAAACAGAGTTGGTGTCTGATCCAAACACACTATGTGACGCTGTAGAAAGCA	----	AspAsnIly	1088
Qy	438	ImePheAsnValLeuILeGLyThrTYrTYrAsnAsnAspCYsAlaAlaLeArg	----	AspAsnIly	456
Db	1089	TCTTATCTTGAAGGAATGAGATATTTGGGAGAAACAGTTGTGATTTGCATTAAACAT	----	AspAsnIly	1148
Qy	457	-LeuILeGLyThrLYsThrCYsArgGLuValTYrGLuPheArgValLYsGLuSerSer	----	AspAsnIly	476
Db	1149	ACTTGGGGGCTTAAAGCGTGTGAGATTTACATTAATTCATGCGCAACAAATCAATG	----	AspAsnIly	1208
Qy	476	e-----	----	AspAsnIly	1147
Db	1209	TACTATGTCATTAGACTTTAACAAACTACAAAGACACATCAGTTACCAAAAAAGT	----	AspAsnIly	1268
Qy	477	eAlaPrOAlaPrOAlaGLuAspValAspThrPrOAsnArgLYsLYsArgLYsAsnArg	----	AspAsnIly	497
Db	1269	ATCTGAAAAGTAGTAGGTGCGTC-----CGCAAAAATGCAAGCTCGC	----	AspAsnIly	1313
Qy	497	GLeUTrPAlaAlaHISCYsArgLYsILeGLuLeuLYsAspGLySerSerAsnHISVa	----	AspAsnIly	517
Db	1314	AAAAATATGCTCGTTATCCGCCCTTTAAAGAAAACAATGATGAGAAAGCTAAGTTT	----	AspAsnIly	1373
Qy	517	LYrAsnLYrGLInPrOCyAsnPhILePrOCrGLInPrOCyAsnSerSerCYsPrOCyVa	----	AspAsnIly	537
Db	1374	TAAAGCATACACACCAATGACT--TGCAAGTCAAAATGTGACACGCAATGCCCTTGTTT	----	AspAsnIly	1430
Qy	537	ILLeAlaGLInAsnPheCYsGLuLYsAspHeCYsGLInCYsSerSerSerGLuCYsGLInAsnArgPh	----	AspAsnIly	557
Db	1431	AACATCAAGAAATTCGTGGAGAAATATTTGGCGGGTCTCAAAAGATTCACAAATCGCTT	----	AspAsnIly	1490
Qy	557	ePrOCrLYsArgCYsLYs--AlaGLInCYsAsnThrLYsGLInCYsPrOCyTYrLeuAl	----	AspAsnIly	576
Db	1491	TGGAGAGATGATTTGTGCAATTTGGCCAATGCACAAATGCACAAATCTCTTTGTTGTCGC	----	AspAsnIly	1555
Qy	576	aValArgGLuCYsAsnPrOCrAsnLYs-----LeuThrCYsGLyAlaAlaAs	----	AspAsnIly	592
Db	1551	TAATGTGATATCGATCCAGATCTTTGTGAGTTGCTCTTACTGTGAGATGGAC	----	AspAsnIly	1610
Qy	592	PhISrTrpAsnSerLYs--AsnValSerCYsLYsAsnCYsSerILeGLInArgLYsLY	----	AspAsnIly	611
Db	1611	TCTTGGTGAACACCAATGCAAATTCATGCAGAAATGCATTCCTCTTCAACCA	----	AspAsnIly	1670
Qy	611	LYsAsnILeLeuAlaPrOsErAsnValaGLyTYrGLyILePheILeLYsAspR	----	AspAsnIly	631

```

Db      1671 TAAAGAATTCCTCATTGGAAAGCTGTAAGTTCATGAGATGGGGTGCATTTACATGGACTC 1730
Qy      631 ova|Gln|yAsn|Glu|Phe|Ile|Ser|Glu|Tyr|Cys|Gly|Glu|Ile|Leu|Ser|Gln|Asp|Glu| 651
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1731 TCTTAATAAAGATGATGATCTCGAGAAATATATACTCGGAGAACTGATCTCATGTATGTAAGC 1790
Qy      651 aaSpA|Arg|Gly|yVal|Tyr|Asp|yTyr|Met|Cys|Ser|Phe|Leu|Phe|Leu|Asn|As 671
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1791 TAAATGAGCGTGGAGAAATGAAAGATTCGGATTGGCTTCTTACCTCTTTTACCTTGAATGA 1850
Qy      671 naSpPhe|Val|Ala|Phe|Ala|Thr|Arg|yGly|yGly|Asn|yIle|Arg|Phe|Ala|Asn|Ile|Ser|Val 691
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1851 TCAGCTCGAATTCATGCTGCTCCGCTTAAGAAAGAAAGAGTTCAAATTTCTCATCTCAGC 1910
Qy      691 IAsn|Pro|Asn|Cys|Tyr|Ala|yVal|Met|Met|Val|Asn|yIle|Asp|Ile|Sar|Gly|Ile|Phe 711
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1911 AAGACCTTAACGTACCGCAAGTGAATGATGTGTAGAGAGAGATCATGAGATTCGTCTAAT 1970
Qy      711 ea|Ala|yArg|Ala|Ile|Gln|Thr|Gly|yGlu|Glu|Phe|Asp|Tyr|Arg|Tyr|Ser|Gln|Al 731
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1971 TCGGAGAGACCAATCGAAGAAAGGTGAGAGACTTTTCTTCGACTACGTACTATGACCA 2030
Qy      731 aaSp|Ala 733
       |||
Db      2031 ACATCGC 2037

RESULT 6
US-09-071-838A-1
: Sequence 1, Application US/09071838A
: Patent No. 7029917
:
GENERAL INFORMATION:
APPLICANT: Fischer, Robert L.
           Onad, Nit
           Kiyosue, Tomohiro
           Yabegari, Ramin
           Margossian, Linda
           Harada, John
           Goldberg, Robert B.
           Goldberg, Robert B.
TITLE OF INVENTION: Nucleic Acids That Control Seed and
                    Fruit Development in Plants
NUMBER OF SEQUENCES: 324
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,838A
FILING DATE: 01-May-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-086100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2136 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

```

LOCATION: 43..2112  
 OTHER INFORMATION: /product= "Arabidopsis  
 fertilization-independent  
 endosperm 1 (FIE1)"  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-071-838A-1

## Alignment Scores:

Pred. No.:	1 77e-60	Length:	2136
Score:	731.50	Matches:	213
Percent Similarity:	42.3%	Conservative:	127
Best Local Similarity:	26.5%	Mismatches:	260
Query Match:	18.1%	Indels:	205
DB:	5	Gaps:	28

US-10-773-302-2 (1-746) x US-09-071-838A-1 (1-2136)

```

Qy 6 LysLysSerGluLysGlyProValCysTPArgLysArgValLysSerGluTyrMetArg 25
Db 17 AGAAAAAGAGAGAGCGAGTGTATGAGAGAGAAAAAGCATGAGACGATGCGAGG 76
Qy 26 Leu--Arg-GlnLeuLysArgPheArgArgAlaAspGluValLysSerMetPheSerSe 44
Db 77 GTTCCACCCGACCTAATCAGATTAAGAGCAATCGA--AAAGAGAGATTCTGCG 133
Qy 44 rAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnIuTPrLysGlnArgAr 64
Db 134 AATATGAGAGAAAT--CGAGCTGAGATAC--TTCCAAGTGTG--CTACTCAT 183
Qy 64 gllGlnProValHisIleLeuThrSerValSerSerLeuArgGly----- 79
Db 184 GCTTACACCATCATCTGTTGACTTAACCGCCGCTGACAGAGATGATATGAGGA 243
Qy 80 -----ThrArgGluCysSerValThrSerAspLeuAspPheProThrG1 94
Db 244 GACAACAATCATTCTTTCGAGAAAGCAAAACCACTTCATTTTCAGTCCCTCATCT 303
Qy 94 nValIleProLeuLysThrLeuAsnAlaVal-----AlaSerValProIleMe 110
Db 304 GATTATTAATCTTACGAGATCAAGGTTATGTTCTTGATGAGATCAAGATTATGCTCT 363
Qy 110 CysSerTrpSerProLeuGlnIleAsnPheMetValGluAspGluThrValIleuHis 130
Db 364 GAGAGAGATGA-CCATTA-----TTCTTGATGAGATGAACCATTTATTCAGAG 413
Qy 130 nIleProTyrMetGlyAspGluValLeuAspGlnAspGlyThrPheIle----- 146
Db 414 TGTCAAGCTTCAATTGTTGAGAGACTACACGATCCATTCATGCGCTTCACCAAAAG 473
Qy 147 -GlnGluLeuIleLysAsnTyrAspGlyLysValHisGlyAspArgGluCysGlyPheI1 166
Db 474 TAGCCAGCTGATGGCTGAAGATGAT--TCTGTGATGTTGTAAGACCAATCTATTATT 530
Qy 166 eAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnIuTyrAsnAspAspAspAs 186
Db 531 GATGCTGAG-----GCATTAAGATTGACAGCAGTGAAGAGATGA 569
Qy 186 pAspAspArgLysAspAspProGlnGlu--ArgGlnGluLysGlnLysAspLeuGln 205
Db 570 GGAAGATGAGAAAGATGAGAGAAATCAAGAAAGAAAGAAATTCGAAATTTTCGAGA 629
Qy 205 pHisArgAspAspLysGluSerArgProProArgLysPheProSerAspLysIlePheG1 225
Db 630 T-----GTAGACCG 638
Qy 225 uAlaIleSerSerMetPheProAspLysGly----- 235
Db 639 ATTATATGACGCGTTGGCAGAGCATATGCTTGGATATCTGCTCGCGCGCTGCT 698
Qy 236 -----ThrAlaGlnGluLeuLysGlnLysGlyTyrLysGlnLeuThrG1 249
Db 699 CGCCAACTACTCGAAGTGGATGTTTCGACATATTGGAAGATGATCAATGAACCTCAAGCT 758

```

```

Qy 249 uGlnGlnLeuProGluAlaLeuProProGluCysThrProAsnIleAspGlyProAsnI1 269
Db 759 TAGAAT-----GATGAACTGCTGG 779
Qy 269 aLysSerValGlnArgGlnSerLeuHisSerPheHisThrLeuPhe----- 285
Db 780 TGAGGCT-----TCTGATTTGACATCCAAAGCAATTAACCTGCTTCCAGATTTTGC 833
Qy 286 -----CysAlaGArgCysPheLysGlyThrSerCysPheLeuHisProPheHisAl 301
Db 834 TGATGAGCTCATTCGCTCGCTGTCATGATATTCATATGTCATATGACAGAACTAT-- 891.
Qy 301 aThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeu-----AspAsnLys 318
Db 892 -GAGCCGAGCTCATGATCCAGAGAAAGCAAACTCTGTTGTTGAGATGAAAGATGACA 950
Qy 318 sProCysGlyProGlnCysTyrGlnHisLeuGlnGluValAlaLysGluPheAlaAlaLe 338
Db 951 ACCATGACAGTGAAGCATTTGTAACCTCAAGAGTGAAGTGAAGAGAA----- 996
Qy 338 uThrAlaGlnArgGlyLysThrProProLysArgProGluGlyAlaGArgArgGlyArgLe 358
Db 997 -----GCTGATCATGTG-----AT 1010
Qy 358 uProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspTh 378
Db 1011 GATATATATATCT----- 1026
Qy 378 rAspSerAspArgLysAlaGlyThrGlnLysGlyLysLeuAsnAspLysGluGluG1 398
Db 1026 ----- 1026
Qy 398 uGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProI1 418
Db 1027 -----AT 1028
Qy 418 eLysMetLysProAsnIleGluProProGluAsnValGluTPrSerGlyAlaGluAlaSe 438
Db 1029 ATCAAAACAAGTTGGTCTGATGATCCAAACAACATATGTGAGAGCTGTGAGAGAGGA 1088
Qy 438 rMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArg----- 456
Db 1089 TCTTACTGAAAGGAATGAGATGATTTTGGAGAAACAGTTGATGTTGATTAACAT 1148
Qy 457 -LeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSerI1 476
Db 1149 ACTTCGGGGCTTAAGAGCTGCTAGAGATTACATTAATGACGCGAACAAGATCAATG 1208
Qy 476 e-----I1 477
Db 1209 TACTATGCTATTAAGCTTTAAACAACATCAACAAGACAAATCAGTTACCAAAAAAGT 1268
Qy 477 eAlaProAlaProAlaGlnAspValAspThrProProArgLysLysLysValGlyHisAl 497
Db 1269 ATCTGAAAAGATGATAGCTGCTG-----CGCAAAAATGAGAGCTGCG 1313
Qy 497 gLeuTrpAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisVa 517
Db 1314 AAAATATCTGTTATCCGCTGCTTAAAGAAAACAATGAGAGAGCTAAAGTTTAA 1373
Qy 517 LTrAsnTyrGlnProCysAspHisAspArgGlnProCysAspSerSerCysProCysVa 537
Db 1374 TAGACATCAACACCATGCACT--TGCAAGTCAAAATGTGACAGCAATGCGCTTGTT 1430
Qy 537 I1leAlaGlnAsnPheCysGlnLysPheCysGlnCysSerSerGluCysGlnAsnArgPh 557
Db 1431 AACTCAGAAAATATGCTCGAGAAATATATGCGGGTCTCAAAAGATTCACAACATCGCTT 1490
Qy 557 eProGlyCysArgCysLys--AlaGlnCysAsnThrLysGlnCysProCysTyrLysuAl 576
Db 1491 TGAGAGATGATATTTGATGCAATTTGCGCAATGACAAATGCAATGATCTCTTGTTCGCTG 1550
Qy 576 aValArgGluCysAspProAspLeuCys-----LeuThrCysGlyAlaAlaAs 592

```

Db 1551 TAATCGTAATGCATCCAGATCTTTGTCGAGTTGCTCTTAGCTGTGAGATGCGAC 1610  
Qy 592 phtitrasperisys---AenValserCyelyasncysSerilegInArgGlySerly 611  
Db 1611 TCTTGTAAGACACCAAGTGCMAATCCAAATCGAAGAACATGCAATTCCTCTTCAACCA 1670  
Qy 611 slyeHiseuleuleuAlaProserAspValAlaGlyTyrpGlyIlePheIlelyAspPr 631  
Db 1671 TAAAGATCTCTCATGTGAAAGTGTGATGTCATGATGGGTGATTAATGAGGACTC 1730  
Qy 631 ovaIGlnlyabangluPheIleSerGlyTyrCyseGlyIleIleSerGlnAspGlnAl 651  
Db 1731 TCTTAAAGAAATGATGTCTCGAGAAATATACAGAACTGATGATCTCTCATGTGAAAG 1790  
Qy 651 aAspArgArgGlylyAspValTyrAspLytyrMetCysSerPheleuPheanleuAsnAs 671  
Db 1791 TAAAGAGCGTGGAGAAATGAAAGATCGGATTTGTTCTTCACTCTTACCTTGAATGA 1850  
Qy 671 naSpPheValValAspAlaThrArgGlyAsnlyIleArgPheAlaAsnHiseSerVa 691  
Db 1851 TCAGCTCAAAATCGATGTGCTGCGCGTAAGAAAGCAAGTTCAATTTCTCAATCACTCAGC 1910  
Qy 691 laSpProAsnCysTyrAlaIlyValMetMetValAsnGlyAspHisArgIleGlyIlePh 711  
Db 1911 AAGACCTTAACCTGCTACGCCCAAGTTGATGATGTGAGAGAGATCAGAGATGTCATTT 1970  
Qy 711 eAlaIlyAspAlaIleGlnThrGlyGlnGluIleuPheAspTyrArgTyrSerGlnAl 731  
Db 1971 TGCGAGAGAGAGCAATCGAAGAGAGAGAGCTTTCTTCACTCTCTATGAGCCAGA 2030  
Qy 731 aAspAla 733  
Db 2031 ACATGCG 2037

RESULT 7  
US-09-061-769A-1  
; Sequence 1, Application US/09061769A  
; Patent No. 6239327  
; GENERAL INFORMATION:  
; APPLICANT: Cold Spring Harbor Labs  
; TITLE OF INVENTION: No. 6239327el Seed Specific Polycomb Group Gene and Methods of  
; FILE REFERENCE: use for Same  
; CURRENT APPLICATION NUMBER: US/09/061,769A  
; PRIOR FILING DATE: 1998-04-16  
; PRIOR APPLICATION NUMBER: 09/061,769  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2315  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: MEAWT  
; NAME/KEY: CDS  
; LOCATION: (14)..(2083)  
US-09-061-769A-1

Alignment Scores:  
Pred. No.: 2,256-60 Length: 2315  
Score: 731.00 Matches: 211  
Percent Similarity: 42.0% Conservative: 120  
Best Local Similarity: 26.7% Mismatches: 247  
Query Match: 18.0% Indels: 243  
DB: 3 Gaps: 28

US-10-773-302-2 (1-746) x US-09-061-769A-1 (1-2315)  
Qy 18 ArgValIlySerGlnTyrMetArgIleuArgGlnIleuIlyAspArgPheArgAlaAspGlu 37  
Db 58 CGAAGTAAATCAATGAT-----AAAAGCAAAATGAAAAGAGAG 96

Qy 38 ValIlySerMetPheSerSerAspAspGlnIlyIleuGlnIlyArgThrGlnIleuAsn 57  
Db 97 ATT-----TCTGATATCAAGAAAAATT-----CGAGCTGAGATCA-TTC 137  
Qy 58 GlnGluThrIlySerGlnArgGlyIleGlnProValHisIleuThrPheSerValSerSerIeu 77  
Db 138 CAAGTGGG---CTACTGATGCTTCACACCATCATGATGCTTTGACTTAAACAGCCCGCTG 194  
Qy 78 ArgGly-----ThrArgGlyCysSerValThrSer 87  
Db 195 CAGAGATGATTAATGAGAGAGACAAATCATCTTTGTGAGAAATGCAAAACCATTC 254  
Qy 88 AspIleuAspPheProThrGlnValIleProLeuIlyThrIleuAlaVal----- 104  
Db 255 GTCAATTCAGTGCCTCATCTGATTAATTTCTTACGAAGATCAAGGTTAATGTTCTTGATG 314  
Qy 105 ---AlaSerValProIleMetTyrSerThrSerProIleuGlnGlnAsnPheMetValGlu 123  
Db 315 AGGATCAAGATTAATGCTTGAAGAGATGTA--CATTA-----TTCTTGATGTA 364  
Qy 124 AspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAspGlnAspGly 143  
Db 365 GATGTACCATTAATTAACAAGTCAAGTCAAGCTTCATTTGTGAGAGCTTACACAGATTCAT 424  
Qy 144 ThrPheIle-----GluGlnIleuIleIlyAsnTyrAspGlyIlyValHisGly 159  
Db 425 ACATGGGTCTTCAACAAAGTCAAGCTGATGGCTGAAAGTAT---TCTGTATTTGT 481  
Qy 160 AspArgGlyCysGlyPheIleAsnAspGluIlePheValGluIleuValAsnAlaIleuGly 179  
Db 482 AAGAGCAAAATCTATTATTGAATGTGAG-----GCATGAA 520  
Qy 180 GlnTyrAsnAspAspAspAspAspAspGlyAspAspProGluGlu--ArgGluGlu 198  
Db 521 TTGAGCAGTGAAGAAATGAGAGAGATGAGAGAGATGAGAGAAATCAAGAAAGAA 580  
Qy 199 LysGlnIlyAspIleuGlnIlyAspHisArgAspAspIlySerArgProProArgIlyPhe 218  
Db 581 AAATGCCAAATTTCTGAGAT----- 601  
Qy 219 ProSerAspIlyIlePheGlnAlaIleSerSerMetPheProAspIlyGly----- 235  
Db 602 -----GTAGACCGATTAATTAATGACGGTGGCCAGGACTATGTTGGATGAT 649  
Qy 236 -----ThraIaGlnIleuIlyGlu 242  
Db 650 CTGATCGGCGCGGCTCTGCCAAGTACCTCGAAGTGAATGTTGCGACATATTGGAA 709  
Qy 243 LysTyrIlyGlnIleuThrGlnGlnIleuProGlyAlaIleuProProGluCysThrPro 262  
Db 710 AGATACATGAATCTCAAGCTTAAGAT----- 736  
Qy 263 AsnIleAspGlyProAsnAlaIlySerValGlnArgGlnIleuIleuIleuIleuIleuIleu 282  
Db 737 -----GATGAACTGCTGGAGGCT-----TCTGATTTGACATCCAAACATAACT 784  
Qy 283 ThrIleuPhe-----CysArgArgCysPheIlyTyrAspCys 294  
Db 785 ACTGCTTCCAGAGATTTTGTGATAGAGCATGATCCGCTCGTGCATGATATTGCAATGT 844  
Qy 295 PheIleuHisProPheHisAlaThrProAsnThrTyrIlyAspGlyAsnThrGlnThrAla 314  
Db 845 CATATGATGAAGAGAT---GAGCCCAAGTCTGATTCACAGCAAGCAAAATCTGATTTG 901  
Qy 315 Leu-----AspAsnIlyProCysGlyProGlnCysTyrGlnIleuGlnIlyAla 331  
Db 902 TTGAGATGAAGATTAACAACATGCAAGTGAATTTTCTTCAAGTGAAGAGTGTG 961  
Qy 332 LysGlnPheAlaAlaIleuThrAlaGluArgIleuIlyThrProProIlyAspProGly 351  
Db 962 ACAGAA-----GCTGATCATGAG----- 979

QY 352 GlyArgArgGlyArgLeuProAsnAsnSerArgProSerThrProThrIleAsn 371  
DB 960 -----ATGATTAATGATTAATCTT----- 997  
QY 372 ValLeuGlnSerIleAspThrAspSerAspArgGluAlaGlyThrGlnThrGlnGlnGln 391  
DB 997 ----- 997  
QY 392 AsnAsnAspIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 411  
DB 997 ----- 997  
QY 412 SerArgGlnThrProIleLeuMetIleProAsnIleGlnProGlnGlnGlnGlnGln 431  
DB 998 -----ATATCAACAAGATTGGTGTCTGATCTCAACAACACTATG 1039  
QY 432 TrpSerGlyAlaGlnAlaSerMetPheArgValIleuIleGlyThrTyTrpAspAsn 451  
DB 1040 TGGAGCGCTGTAGAGAGATCTTTACTTGAAGAATTGAGATATTGGGAGAAACAGT 1099  
QY 452 CysAlaIleAlaArg-----LeuIleGlyThrTyTrpCysArgGlnValTyTrp 469  
DB 1100 TGTGATGTGCTTAAACATCTTGGGGGCTTAAAGCGTCTAGAGATTTCATATAC 1159  
QY 470 ArgValIleGlnSerSerIle----- 476  
DB 1160 ATGCCGGAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1219  
QY 477 -----IleAlaProAlaProAlaGlnAspValAspThrProArg 490  
DB 1220 AATCAGTTACCAAAAAGTATCTGAAAAGTATAGTGGTCTC----- 1264  
QY 491 LysIleValArgValIleArgLeuThrAlaAlaIleCysValGlyIleGlnLeuVal 510  
DB 1265 CGCAAAAATCGAGCTCCGAAATAATGCTCTGATCCGCTCTTAAAGCAAACT 1324  
QY 511 AspGlnSerSerAsnIleValIleValIleValIleValIleValIleValIleVal 530  
DB 1335 AGTGGAGAGCTAAGTTTATTAAGCACTACACACATGACT--TGCAGTCAAAAGT 1381  
QY 531 AspSerSerCysProCysValIleAlaGlnAsnAspCysGlnValPheCysGlnCys 550  
DB 1382 GGAACGACATGCGCTTGTTTAACTGACGAAATTCCTGCGAATATATGGCGGCTCA 1441  
QY 551 SerGlnCysGlnAsnAspPheProGlnCysArgCysVal--AlaGlnCysAsnThr 569  
DB 1442 AAGGATGCAACAATCGCTTGAAGATGATATGTCGAATGCGCAATGCAACAATGCA 1501  
QY 570 GlnCysProCysGlnLeuAlaValAlaArgGlnCysAspProAspLeuCys----- 585  
DB 1502 CAATGTCTGTGTGCTGCTAATGTCGATGTCGATGTCGATGTCGATGTCGATGTCCT 1561  
QY 586 LeuThrCysGlyAlaAlaAspIleTrpAspSerIle--AsnValSerCysValAsnCys 604  
DB 1582 CTGACTGTGAGATGAGCACTTGTGAGACACAGTGCATTCATGCAAGAACATG 1621  
QY 605 SerIleGlnArgGlySerIleValIleValIleValIleValIleValIleValIle 624  
DB 1622 CAATTCCTCTTCAACAATTAAGATTCCTCATTTGGAAGTCTGATGTTCTGAGATG 1681  
QY 625 GlyIlePheIleValAspProValGlnIleValAsnGlnIlePheIleSerGlnTyTrp 644  
DB 1682 GGTGCAATTACATGGGACTCTTTAAAGAAATGAGTATCTCGAGAAATATATCTGAGA 1741  
QY 645 IleIleSerGlnAspGlnAlaAspIleArgValIleValIleValIleValIleVal 664  
DB 1742 CTGATCATCTGATGAGATGATGAGCTGTGAGAAATTAAGAGATCGATTTGTTCTTC 1801  
QY 665 PheLeuPheAsnIleValAsnAspPheValAlaAspAlaThrArgValGlyValAsn 684  
DB 1802 TACCTCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1861  
QY 685 ArgPheAlaAsnIleSerValAsnProAsnCysTyTrpAlaValIleMetValAsnGly 704

DB 1862 AAATTTTCATCATCTGACGAAAGCTTAACGTCTACGCGCAAGTTGATGTTGAGAGCA 1921  
QY 705 AspHisArgIleGlyIlePheAlaValArgAlaIleGlnThrGlnGlnLeuPhePhe 724  
DB 1922 GATCAGAGAAATTTGGTCTATTGTCGAGAGAGCAATCGAAGAAAGTGAAGAGCTTTCTTC 1981  
QY 725 AspTyArgTyTrpSerGlnAlaAspAla 733  
DB 1982 GACTACTGCTATGACCAAAACATGCG 2008  
RESULT 8  
US-10-176-884-16  
Sequence 16, Application US/10176884  
Patent No. 6906244  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert  
APPLICANT: Kinoshita, Tetsu  
APPLICANT: Yabegaki, Ramin  
APPLICANT: Gehring, Mary  
APPLICANT: Okamoto, Jack  
APPLICANT: Dang, Van-Dinh  
APPLICANT: The Regents of the University of California  
APPLICANT: Ceres, Inc.  
TITLE OF INVENTION: Compositions and Methods for Modulating Plant  
TITLE OF INVENTION: Development  
CURRENT REFERENCE: 023070-116710US  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/300,506  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 16  
LENGTH: 2799  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: MEA-15G-BSA1  
OTHER INFORMATION: chimeric HAT nucleic acid construct encoded by  
OTHER INFORMATION: transgene plasmid pCR304-5UNS MEA-15G-BSA1  
US-10-176-884-16  
Alignment Scores:  
Pred. No.: 3,066-60 Length: 2799  
Score: 731.00 Matches: 211  
Percent Similarity: 42.0% Conservative: 120  
Best Local Similarity: 26.7% Mismatches: 247  
Query Match: 18.0% Indels: 213  
DB: 3 Gaps: 28  
US-10-773-302-2 (1-746) x US-10-176-884-16 (1-2799)  
QY 18 ArgValIleSerSerGlnTyTrpMetArgLeuArgGlnIleuValArgPheArgAlaAspGln 37  
DB 45 CGAATTAATCAAGT-----AAAGGCAAAATCGAAAAGAG 83  
QY 38 ValIleSerMetPheSerAsnArgGlnIleValIleLeuGlnArgGlnIleValAsn 57  
DB 84 ATT-----TCTGCAATCAAGAGAAATT-----CGAGCTGAGATACA-TTC 124  
QY 58 GlnGlnTyTrpGlnArgGlnIleGlnProValIleIleuThrSerValSerSerIleu 77  
DB 125 CAAGTGTG---CTACTCATCTTCAACATCAATCGTTGACTTAAACAGCCCGCTG 181  
QY 78 ArgGly-----ThrArgGlnCysSerValThrSer 87  
DB 182 CAGAGATGATTAAGAGAGACAAATGACTTTTGTGAGAAATGAGAAACCATTC 241  
QY 88 AspLeuAspPheProThrGlnValIleProLeuValIleProLeuValIleProLeuVal 104  
DB 242 GTCAATTCAGGCGCTCATCTGATTAATTTTGAAGAGATCAAGGTTATGTTCTTGATG 301

OY	105	---	AlaSerValProIleMetTyrSerTrpSerProLeuGlnIleAsnProMetValGlu	123	
Db	302	AGGATCAAGATTTAATGCTCTTGAAAGAGATGTA--CCATTA-----	TTTCTGATGAA	351	
OY	124	AspGluThrValLeuHisAsnIleProTyrMetGluArgGluValLeuAspGlnAargGly	143		
Db	352	GATGACCATATTATTTCCAAAGTCAGTCAAGCTTCCAAATTGTGTGAAGAGCTACACAGATCCAT	411		
OY	144	ThrPheIle-----	GluGluLeuIleLysAsnTyrAargGlyLysValHisGly	159	
Db	412	ACATGGGCTTCACACCAAAAGTACGACAGCTGATGCTGAAAGTATGAT--	TCGTGATTTGGT	468	
OY	160	AspArgGluCysGlyPheIleAsnAargGluIlePheValGluLeuValAsnAlaLeuGly	179		
Db	469	AAGAGCAAACTATATTATTGAATGGTAG-----	GCACTTGAA	507	
OY	180	GlnTyrAsnAargAargAargAargAargAargGlyAargAargProGluGlu--	ArgGluGlu	198	
Db	508	TTGACGAGTACAGAGATGAGAGACATGAAAGAAAGATGAGAGAAATTCAGAAAGAA	567		
OY	199	LysGlnLysAerLeuGluAerHisIaArgAargLysGluSerArgProAargLysPhe	218		
Db	568	AAATGCCAATTTTCTGAAGAT-----		588	
OY	219	ProSerAerLysIlePheGlnAlaIleSerMetPheProAargLysGly-----		235	
Db	589	-----	GTAACCGAATTTATATGACAGGTTGGGACAGACTTGGTTTGATGAT	636	
OY	236	-----	ThrAlaGluLeuLysGlu	242	
Db	637	CTGGTCGGCGGCGGCTCTGCGCAAGTACTCCAAAGTGATGTTTGGACATATTTGAA	696		
OY	243	LysTyrLysGluLeuThrGluGlnIleuProGluAlaLeuProGluCysThrPro	262		
Db	697	AGATGCATGAACTCAAGCTTAAGAT-----		723	
OY	263	AsnIleAerGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSerPheHis	282		
Db	724	-----	GATGAACTGCTGTGAGGCT-----	TCTGATTTGACCTCCAAACAAATACT	771
OY	283	ThrIleuPhe-----	CysArgCysPheLysTyrAspCys	294	
Db	772	ACTGCTTCCAGAGATTTTGCATAGACTCATTCGCGCTGCTGATGATTTGGATTTGT	831		
OY	295	PheLeuIleProPheIleAlaThrProAsnThrTyrLysArgLysAsnThrGluThrAla	314		
Db	832	CATATGCATGAGAACTAT--GAGCCGAGCTATAGTCCAGCAAAACAAATCATGTTTG	888		
OY	315	Leu-----	AsrAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluValA	331	
Db	889	TTTGAGAGTGAAGATAGCAACACATGACAGTGCATTTGTTCCTCCAAAGTCAAGAGTGTG	948		
OY	332	LysGluPheAlaAlaLeuThrIleGluArgIleLysThrProProLysArgProGly	351		
Db	949	ACAGAA-----	GCTATCATGTG-----	966	
OY	352	GlyArgArgGlyArgLeuProAsnAsnSerArgProSerThrProThrIleAsn	371		
Db	967	-----	ATGATATATATTAATCT-----	984	
OY	372	ValLeuGluSerLysAerThrAerSerAerArgGluAlaGlyThrGluThrGlyGly	391		
Db	984	-----		984	
OY	392	AsnAsnAerLysGluGluGluGluLysLysAerGluThrSerSerSerSerGluAlaAsn	411		
Db	984	-----		984	
OY	412	SerArgCysGlnThrProIleLysMetLysProAsnIleGluProProGluAsnValGlu	431		
Db	985	-----	ATATCAAAACAAAGATTGGTGCTCTAGATCCAAACAAACATATG	1026	
OY	432	TrpSerGlyAlaGluAlaSerMetPheArgValLeuIleGlyThrTyrTrpAspAsn	451		

[illegible]

TITLE OF INVENTION: No. 6239327e1 Seed Specific Polycomb Group Gene and Methods of  
 TITLE OF INVENTION: use for Same  
 FILE REFERENCE: mea genes  
 CURRENT APPLICATION NUMBER: US/09/061,769A  
 CURRENT FILING DATE: 1998-04-16  
 PRIOR APPLICATION NUMBER: 09/061,769  
 PRIOR FILING DATE: 1998-04-16  
 NUMBER OF SEQ ID NOS: 21  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO 3  
 LENGTH: 2327  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 NAME/KEY: Variation  
 LOCATION: (1760)..(1761)  
 US-09-061-769A-3

## Alignment Scores:

Pred. No.:	1,69e-59	Length:	2327
Score:	722.00	Matches:	212
Percent Similarity:	41.7%	Conservative:	119
Best Local Similarity:	26.7%	Mismatches:	247
Query Match:	17.8%	Indels:	217
DB:	3	Gaps:	29

US-10-773-302-2 (1-746) x US-09-061-769A-3 (1-2327)

QY ArgVallyserGluTyrMetArgLeuArgGlnLeuYsArgPheArgGlnAlaSerGlu 37  
 DB CGAACTAAATCGAT-----AAAAGCGAAATCGAAAGAGAG 96  
 QY 38 VallySerMetPheSerSerAnaArgGlnYsIleLeuGlnArgThrGlnIleLeuAsn 57  
 DB ATT-----TCTGCATATCAAGAGAAATT-----CGAGCTGAGATACA-TTC 137  
 QY 58 GlnGlnTrrlysglnArgArgIleGlnProValHsIleLeuThrSerValSerSerLeu 77  
 DB 138 CAAGTGTG--CTACTCATGCTTCACACCATCAATCGTTTGAATTAACCGCCGCTG 194  
 QY 78 ArgGly-----ThrArgGluCysSerValThrSer 87  
 DB 195 CAGAGGATGATATGAGAGAGCAACAATCATCTTTGTCGAGAAATGCAAAACCACTTC 254  
 QY 88 AspleuAspPheProThrGlnValIleProLeuYsThrLeuAsnAlaVal----- 104  
 DB 255 GTCATTTGAGTCCCTCATCTGATATTAATTTTACGAGATCAAGGTTATGTTCTTGATG 314  
 QY 105 ---AlaSerValProIleMetYsSerTrpSerProLeuGlnGlnAsnPheMetValGlu 123  
 DB 315 AGGATCAAGATTATGCTCTTGAAAGAGATGTA-CCATTA-----TTCTTGTATGTA 364  
 QY 124 AspGlnThrValLeuHsIleAsnIleProTyrMetGlnYsArgGlnValIleAspGlnAspGly 143  
 DB 365 GATGACCATTTATTCACAGTGTCAAGCTTCCAATTTGTGAGAACTGACCAAGATCCATT 424  
 QY 144 ThrPheIle-----GlnGlnLeuIleYsAsnTyrAspGlyValValHsIleGly 159  
 DB 425 ACATGGGTCTTCACCAAAAGTAGCCAGTGAATGGGTGAAGATGAT---TCTGTGATGGT 481  
 QY 160 AspArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGly 179  
 DB 482 AAGACAAATCTATTATTGTAATGGTAG-----GCACCTAGAA 520  
 QY 180 GlnTyrAsnAspAspAspAspAspAspAspGlnYsAspArgProGlnGlu---ArgGlnGlu 198  
 DB 521 TTGACAGTGAAGAGATGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATCAAGAGAA 580  
 QY 199 LysGlnYsAspLeuGlnAspHsIleArgAspAspGlnYsArgProArgGlnYsPhe 218  
 DB 581 AAATCGAATTTTCTGAAGAT----- 601

QY 219 ProSerAspYsIlePheGlnAlaIleSerSerMetPheProAspYsGly----- 235  
 DB 602 -----GTAGACCAATTTATATGACCGGTGGGACGACATATGTTGATGAT 649  
 QY 236 -----ThraIleGlnIleuLeuYsGlu 242  
 DB 650 CTGTCGTGCGGCGGTCTCTGCGCAAGTACCTCGAAGTGAATGTTTCCGACATATTTGGA 709  
 QY 243 LysTyrYsGlnLeuThrGlnGlnIleuProGlnYsAlaLeuProProGlnCysThrPro 262  
 DB 710 AGATCAATGAACTCAAGCTTAAGAT----- 736  
 QY 263 AsnIleAspGlyProAsnAlaYsSerValGlnArgGlnSerLeuHsIleSerPheHs 282  
 DB 737 -----GATGCACTGCTGTGAGGCT-----TGTGATTCACATCCAAAGCAATTA 784  
 QY 283 ThrLeuPhe-----CysArgArgCysPheYsYsArgCys 294  
 DB 785 ACTGCTTCCAGGATTTTGTGATGAGAGTCAATGCGCTGTCGTCATGATATTCGATG 844  
 QY 295 PheLeuHsIleProPheHsIleAlaThrProAsnThrTyrYsValGlyAsnThrGlnThrAla 314  
 DB 845 CATATGATGAGAAAGTAT---GAGCCCAAGCTTACATCCAGGAGACAAATCTTGGTTG 901  
 QY 315 Leu-----AspAsnYsProCysGlyProGlnCysTyrGlnHsIleuGlnGlyAla 331  
 DB 902 TTGAGATGAGAGATGAGAGACATGACATGACATTTGATCTCAAGGTGAGAGTGTG 961  
 QY 332 LysGlnPheAlaAlaAlaLeuThrAlaGlnArgIleYsThrProProYsAspProGly 351  
 DB 962 ACAGAA-----GCTGATCATGTG----- 979  
 QY 352 GlnArgArgGlyValArgLeuProAsnAspSerSerArgProSerThrProThrIleAsn 371  
 DB 980 -----ATGATATGATATCACTT----- 997  
 QY 372 ValLeuGlnSerYsAspThrAspSerAspArgGlnAlaGlyThrGlnYsGlyGlu 391  
 DB 997 ----- 997  
 QY 392 AsnAsnAspYsGlnGlnGlnGlnYsYsAspGlnThrSerSerSerGlnAlaAsn 411  
 DB 997 ----- 997  
 QY 412 SerArgCysGlnThrProIleYsMetYsProAsnIleGluProProGlnAsnValGlu 431  
 DB 998 -----ATATCAAAACAAGATTGTGTGTCAGATCCAAACAACAACATG 1039  
 QY 432 TrpSerGlnAlaGlnAlaSerMetPheArgValLeuIleGlyThrTyrTrpAspAsnHe 451  
 DB 1040 TGAACGCTGTAGAGAGAGATCTTACTTGAAGAGATTTGAATTTGGAGAAACAGT 1099  
 QY 452 CysAlaIleAlaAsp-----LeuIleGlyThrYsThrCysArgGlnValYsGlnPhe 469  
 DB 1100 TGTGATGTGATTAACATATCTTCCGGGGCTTAAGAGTGTCTTGAAGATTTACATTAAC 1159  
 QY 470 ArgVallysglnSerSerIle----- 476  
 DB 1160 ATGCGCAACAAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1219  
 QY 477 -----IleAlaProAlaProAlaGlnIleuAspYsValAspThrProProArg 490  
 DB 1220 AATCAGTTACCAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1264  
 QY 491 LysYs 510  
 DB 1265 CGCAAAAATGAGAGTCCGAAATATGCTGTTATCCGCTGTTTAAAGAAAACAAT 1324  
 QY 511 AspGlySerSerAsnHsIleValTyrAsnTyrGlnProCysAspAsnHsIleProArg 530  
 DB 1325 AGTGAGAGAGCTAATTTATATGACATACACACATGACAT---TGCAGATCAAAATGT 1381  
 QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnYsPheCysGlnCysSer 550



DB 1382 GGAACGAAATGCCCTTGTTAACATCAGAAATATGTCGGAATATATGCGGCTCA 1441  
QY 551 SerGluCyseGlnaAspPheProGlyCyseArgCyseLys---AlaGlnCyseAsnThrLys 569  
DB 1442 AAGATTGGCAACAATCGCTTTGGAGATGATATGTCATATGTCATTCGCAATGCA 1501  
QY 570 GlnCyseProCyseTyrLeuAlaValArgGluCyseAspProAspLeuCyse----- 585  
DB 1502 CAATGCTCTTGTGCTGCTATATCGTAATGCCATCCAGATCTTTGTTCGAGTTGCT 1561  
QY 586 LeuThrCyseGlyAlaAlaAspHisTyrAspSerLys---AsnValSerCyseLysAsnCyse 604  
DB 1562 CTGAGCTGTGAGATGCGCACTCTTGAGACACCAGTCGAATTCATGCAAGAACATG 1621  
QY 605 SerIleGlnArgGlySerLysLeuHisIleLeuLeuLeuAlaProSerAspValAlaGlyTyr 624  
DB 1622 CAATTCCTCTTCAACCAATTAAGATGTCATGGAAGCTGATGATTCATGATGCG 1681  
QY 625 GlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGluTyrCyseGlyGlu 644  
DB 1682 GGTGATTACATGAGATGCTCTTAAAGATGATATCTCGAGATATATCGAGAA 1741  
QY 645 IleIle-----SerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
DB 1742 CTGATCACTCACTCATGATCATGATGATGATGATGATGATGATGATGATGATGATG 1801  
QY 661 TyrMetCyseSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgLys 680  
DB 1802 ATGTGTTCTTCTTACCTCTTTCCTGATGATGATGATGATGATGATGATGATGAT 1861  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCyseTyrAlaLysValMet 700  
DB 1862 GGAAGAGAGTTCAATTTCTCAATCACTCAGCAAGACCTTAACCTGTAAGCCAAAGTTGATG 1921  
QY 701 MetValaAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
DB 1922 ATTTGAGAGGAGATCAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1981  
QY 721 GlnLeuPhePheAspTyrArgTyrSerGlnIleAspAla 733  
DB 1982 GAGCTTTCTTCTGACTACTGATGACCAAGAACATGCG 2020

RESULT 10  
US-09-061-769A-4  
; Sequence 4, Application US/09061769A  
; Patent No. 6239327  
; GENERAL INFORMATION:  
; APPLICANT: Cold Spring Harbor Labs  
; TITLE OF INVENTION: No. 623932761 Seed Specific Polycomb Group Gene and Methods of  
; TITLE OF INVENTION: use for Same  
; FILE REFERENCE: mea genes  
; CURRENT APPLICATION NUMBER: US/09/061,769A  
; CURRENT FILING DATE: 1998-04-16  
; PRIOR APPLICATION NUMBER: 09/061,769  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 2322  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: MEA2  
; NAME/KEY: variation  
; LOCATION: (1756)..(1763)  
US-09-061-769A-4

Alignment Scores:  
Pred. No.: 4.78e-58 Length: 2322  
Score: 707.00 Matches: 211  
Percent Similarity: 41.8% Conservative: 120  
Best Local Similarity: 26.6% Mismatches: 247

Query Match: 17.5% Indels: 216  
DB: 3 Gaps: 29  
US-10-773-302-2 (1-746) x US-09-061-769A-4 (1-2322)

QY 18 ArgValLysSerGluTyrMetArgLeuArgGlnLeuLysAspPheArgAlaAspGlu 37  
DB 58 CGAATCAATCAAT-----AAAAGACCAATCGAAAAGAGAG 96  
QY 38 ValLysSerMetPheSerSerAsnArgGlnLysIleLeuGluLysThrGlnIleLeuAsn 57  
DB 97 ATT-----TTCGATATCAAGAGAAAT-----CGAGCTGAGATACA-TTC 137  
QY 58 GlnGluTyrLysGlnArgAlaGlnIleGlnProValHisIleLeuThrSerValSerSerLeu 77  
DB 138 CAAGTGTG---CTACTCATGCTTCAACCATCATCATGTTGTTCACTTAACCCGCGCTG 194  
QY 78 ArgGly-----ThrArgLysCyseSerValThrSer 87  
DB 195 CAGAGATGATTAATGAGAGACAAATCACTTTGTGAGATGCAAAACCCACTTC 254  
QY 88 AspLeuAspPheProThrGlnValIleProLeuLysThrLeuAsnAlaVal----- 104  
DB 255 GTGATTTCAAGTCCCTCATGATTAATTTCTTAAGAGATCAAGGTTATGTTCTGATG 314  
QY 105 ---AlaSerValProIleMetTyrSerTyrSerProLeuGlnGlnAsnPheMetValGlu 123  
DB 315 AGGATCAAGATTATGCTCTTGAAGAGATGTA-CCATTA-----TTTCTTGATGAA 364  
QY 124 AspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValIleAspGlnAspGly 143  
DB 365 GATGTACCATTAATTAACAAGTGCACCTTCCAAATTTGAGAGCTTACCAAGCATTCAT 424  
QY 144 ThrPheIle-----GlnGluLeuIleLysAsnTyrAspGlyLysValHisGly 159  
DB 425 ACATGGGCTTCAACCAAAAGTACGACGTAGTGCAGAAAGTAT---TCGTGATTTGCT 481  
QY 160 AspArgGluCyseGlyPheIleAsnAspGluIlePheValGlnLeuValAsnAlaLeuGly 179  
DB 482 AAGAGACAAATCTATTGTAATGCTGAG-----GCACTTGAA 520  
QY 180 GlnTyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 198  
DB 521 TTGAGCAGTGAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 580  
QY 199 LysGlnLysAspLeuAspHisArgAspAspLysGluSerArgProProArgLysPhe 218  
DB 581 AATGCGAATTTCTGAAGAT----- 601  
QY 219 ProSerAspLysIlePheGlnAlaIleSerSerMetPheProAspLysGly----- 235  
DB 602 -----GTAGACCGATTATATGACGCGTGGCGAGACTATGTTGGATGAT 649  
QY 236 -----ThraIaGlnLysLeuLysGlu 242  
DB 650 CTGCTGCGCGCGCTGCTGCGCAAGTACTCGAAGTGAATGTTTGGACATATTTGGA 709  
QY 243 LysTyrLysGlnLeuThrGlnGlnLeuProGluAlaLeuProGluCyseThrPro 262  
DB 710 AGATTCATGAACTCAAGCTTAAGAT----- 736  
QY 263 AsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSerPheHis 282  
DB 737 -----GATGGAATGCTGTGAGGCT-----TCTGATTTGACATCCAGACATTA 784  
QY 283 ThrLeuPhe-----CysArgArgCysePheLysTyrAspCyse 294  
DB 785 ACTGCTTTCCAGATTTTGTGATAGAGCTATGCGCTGCTGATGATGATGATGATGATG 844  
QY 295 PheLeuHisProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAla 314  
DB 845 CATATGATGAGAGATAT---GAGCCGAGCTGATGATCCAGCAAGCAAAATTAAGTTTG 901

```

QY 315 Leu-----AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluGlyAla 331
DB 902 TTGAGAGTGAAGATGACACACATTCAGTGCATTTTACTCAAGGTGAGAGTGG 961
QY 332 LysGluPheAlaAlaLeuThrAlaGluArgIleTyrThrProProLysArgProGly 351
DB 962 ACAGAA-----GCTGATCATGTC----- 979
QY 352 GlyArgArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsn 371
DB 980 -----ATGATGATATGATTAATCTCT----- 997
QY 372 ValLeuGluSerLysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyGlu 391
DB 997 ----- 997
QY 392 AsnAsnAspLysGluGluGluGluLysLysAspGluThrSerSerSerSerGluIleAsn 411
DB 997 ----- 997
QY 412 SerArgCysGlnThrProIleLysMetLysProAsnIleGluProProGluAsnValGlu 431
DB 998 -----ATATCAACAAGATTTGTGCTCAGATCCAAACAACACTATG 1039
QY 432 TPSSerGlyAlaGluAlaSerMetPheArgValIleuIleGlyThrTyrTyrAspAsn 451
DB 1040 TGGAGCGCTGTAGAGAGAGATCTTTAAGAGATTTGAAGATTTGAGAGAAACACT 1099
QY 452 CysAlaIleAlaArg-----LeuIleGlyThrLysThrCysArgGlnAlaIleTyrGluPhe 469
DB 1100 TGTGATGTTGATTAACATCTTGTGGGGCTTTAGAGCTGCTAGAGTTTACATTAAC 1159
QY 470 ArgValLysGluSerSerIle----- 476
DB 1160 ATGCCGCAACAGATCAATGTATGTATGATAGACTTAAACAACACTACAGAAAGACAC 1219
QY 477 -----IleAlaProAlaProAlaGluAspValAspThrProProArg 490
DB 1220 AATCAGGTTTACCAAAAAGTATCTGAAAAGTATGAGTGGCTG----- 1264
QY 491 LysLysLysArgLysArgLysArgLeuTyrAlaAlaHisCysArgLysIleGlnLeuLysLys 510
DB 1265 CGCAAAAATCGAGACTCGCAAAAATATGCTGCTATCCGCCCTTTAAAGAAACACT 1324
QY 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnProCys 530
DB 1325 AGTGGAGAAAGTAAATTTTATTAAGCACTACACACATGAGACT---TGCAGTCAAAATGT 1381
QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnLysPheCysGlnCysSer 550
DB 1382 GGACAGCAATGCCCTTTGTTTAATCGACGAAAATTCCTCGAGAAAATATGTCGGGTGCTCA 1441
QY 551 SerGluCysGlnAspArgPheProGlyCysArgCysLys---AlaGlnCysAsnThrLys 569
DB 1442 AAGATTTGCAACAATCGCTTTGAGAGATTAATTTGCAATTTGGCCAAATGCAACAATCA 1501
QY 570 GlnCysProCysTyrIleuAlaValIleArgGlnCysAspProAspLysCys----- 585
DB 1502 CAATGCTCTGTTTGTCTGCTATGATGATGATCCGATCCGATCTTTGTCGAGATTGCTCT 1561
QY 586 LeuThrCysGlyAlaAlaAspHisThrAspSerLys---AsnValSerCysLysAsnCys 604
DB 1562 CTTACCTGTGAGATGCGACTCTTGATGACACACAGTGCAAATCCAAATGCAAGAACATG 1621
QY 605 SerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAspValAlaIleTyr 624
DB 1622 CAATTCCTCTTCAACCAACAAATGAATTCATTCATTTGAAACTCTGATCTTATGATGATG 1681
QY 625 GlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGluTyrCysGlyGlu 644
DB 1682 GGTGATTTTACATGCGACTCTCTTAAAGAAATGATGATCTCGAGAAATATACTCGAGAA 1741
QY 645 IleIle-----SerGlnAspGluAlaAspArgArgGlyLysValTyrAspLysTyrMe 662

```

```

DB 1742 CTGATCACTCATGATCATGATGATGAGCTAATAGCGGTGGAGAAATGAGATCGGATTGG 1801
QY 662 tCysSerPheLeuPheAsnLeuAsnAspPheValAlaIleAlaIleAlaIleAlaIle 682
DB 1802 TTCTTCTTACCTCTTACCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1861
QY 682 nLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMetMetVal 702
DB 1862 CGAGTTCAAAATTTCTCAATCACTCGCAAGACTTACGCTACGCGCAAGTTGATATGATGT 1921
QY 702 lAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGluGlu 722
DB 1922 GAGAGAGATCGAGAGATTTGCTCTTTTCCGAGAGAGACATCGAAGAGGAGAGACT 1981
QY 722 uPhePheAspTyrArgTyrSerGlnAlaAspAla 733
DB 1982 TTCTTCTGACTTACTGCTATGAGCAGAACATGCG 2015

RESULT 11
US-09-699-266A-10
/ Sequence 10, Application US/09699266A
/ Patent No. 6559354
/ GENERAL INFORMATION:
/ APPLICANT: Ma, Hongchang
/ APPLICANT: Morakinyo, Layo O.
/ APPLICANT: Odell, Joan T.
/ APPLICANT: Orosco Jr., Emil M.
/ APPLICANT: Rafaliski, J. Antoni
/ TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS
/ FILE REFERENCE: B01164 US NA
/ CURRENT APPLICATION NUMBER: US/09/699,266A
/ PRIOR FILING DATE: 2000-10-27
/ PRIOR APPLICATION NUMBER: PCT/US99/08385
/ PRIOR FILING DATE: 1999-04-16
/ PRIOR APPLICATION NUMBER: 60/083,212
/ PRIOR FILING DATE: 1998-04-27
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 1406
/ TYPE: DNA
/ ORGANISM: Trifolium aestivum
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (513)
/ OTHER INFORMATION: n=a, c, g, or t
US-09-699-266A-10

Alignment Scores:
Pred. No.: 1.18e-53 Length: 1406
Score: 658.00 Matches: 139
Percent Similarity: 56.04 Conservative: 49
Best Local Similarity: 41.44 Mismatches: 113
Query Match: 16.28 Indels: 35
DB: Gaps: 6

US-10-773-302-2 (1-746) x US-09-699-266A-10 (1-1406)
QY 432 TPSSerGlyAlaGluAlaSerMetPheArgValIleuIleGlyThrTyrTyrAspAsn 451
DB 10 TGGATACCTTGAAGAGAGATTTATATCTGAAGAGAAATGAGATTTTGGAAAAATATG 69
QY 452 CysAlaIleAlaArg-----LeuIleGlyThrLysThrCysArgGlnVal-----Tyr 467
DB 70 TGTCTAATATGACAGAAACCTTATATGTCGCGTGAAGAACTCTCAGAGAGAGTGGCTGATC 129
QY 468 GluPheArgValLysGluSerSerIleAlaProAlaProAlaGluAspValAspThr 487
DB 130 ATGACACAAATGTCGAGCAAAAGATGATTAATTCATTTGGCGGATTTCAAGAAACT 189
QY 488 ProPro----- 489

```



```

OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1
FEATURE:
NAME/KEY: CDS
LOCATION: (2)..(7015)
OTHER INFORMATION: fertilization-independent endosperm 1 (PIE1)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(7013)
OTHER INFORMATION: fertilization-independent endosperm 1 (PIE1)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3
US-09-177-249-6

Alignment Scores:
Pred. No.: 5,12e-31 Length: 7015
Score: 436.00 Matches: 122
Percent Similarity: 34.5% Conservative: 42
Best Local Similarity: 25.6% Mismatches: 78
Query Match: 10.8% Indels: 235
DB: 3 Gaps: 11

US-10-773-302-2 (1-746) x US-09-177-249-6 (1-7015)
QY 491 LysLysLysArgLysHisArgLeuThrPalaIaHisCysArgLysIleGlnLeuLysLys 510
   |||||
DB 4782 CGCAAAAATCGAGACTCGCAAAATATGCTCTTCGCGCTTTAAAGAAAACACT 4841
   |||||
QY 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnProCys 530
   |||||
DB 4842 AGTGAGAGAGCTAAGTTTATTAAGCACTACACCATGCACT--TGCAAGCAAAATGT 4898
   |||||
QY 531 AspSerSerCysProCysValIleAlaGlnAsnProCysGlnLysPheCys----- 547
   |||||
DB 4899 GGACAGCAATGCCCTTGTTTACTGACGAAATCTCGAGAAATATTTGCGGATATGTC 4958
   |||||
QY 547 ----- 547
DB 4959 ATTCAATTTTCTTAAGCCGGAAGATCCATGAGATTTAATTGAACATGAGTTGTATTT 5018
   |||||
QY 548 -----Gln-CysSerSerGluCysGlnAsnArgPheProGlyCysArgCysLys--Al 564
   |||||
DB 5019 TTGTTTCAAGTCTCAAGAGATTGCAACATGCTTTGGAGAGATGTAATGTGCAATGG 5078
   |||||
QY 564 aglncysasnthrlcysglnCysProCysGlyLeuAlaValAlaGlnLysCysAspProAspLe 584
   |||||
DB 5079 CCAATGCACAATTCGACAAATGCTTGTGCTCTTAATCGTAATGCGATCCAGATCT 5138
   |||||
QY 584 uCysLeuThrCysGlyAlaIleAspHisGlyPheAsp----- 595
   |||||
DB 5139 TTGTCGAGTGTGCTCTTGAAGTAACTTTCATCTTCAATATCTTTATACAAATTTCTA 5198
   |||||
QY 596 -----SerLysAsn----- 598
DB 5199 TAATCAAGTAACTTCAACCAAAAGTCTTATAAAAAAATTATATATAGCTGAGAGA 5258
   |||||
QY 599 -----ValSerCysLysAsn-CysSer----- 605
DB 5259 TGGCACTCTGCTGAGACACAGTCGCAATTCGAATGCAAGAACAGCATTCCTCTTCA 5318
   |||||
QY 606 -----IleGlnArgLysSerLysLysHis----- 613
DB 5319 AACCAATAAAAGGTAATCAACGTAACCAATCCGTACCGAAATTTAAAATAATTAACGA 5378
   |||||
QY 614 -----LeuLeuLeuAlaProSerAspValA 622
DB 5379 AAGACATTTAACTATCATTTCCGTAATTTTACTAATTTCTATGGAAGTGTGATGTTT 5438
   |||||
QY 622 laGlyTyrGlyIlePheIle----- 628
DB 5439 ATGATGGGGTGCACTTATCATGCTGTAGCAATCATGTAAATATAGAAATAGTTTAATAG 5498
   |||||
QY 628 ----- 628

```

```

DB 5499 TTATGTGCACTTATTAACACTTTTTTTTTTAAATATATGTTTATTAAGTACCATTA 5558
QY 629 -----LysAspProValGlnLysAsnGlnPheIleS 639
DB 5559 AATATATTTGTGTATATGTTTGAACCCGTCAGACTCTCTTAAAAAGATGATCTTCG 5618
   |||||
QY 639 engLutyrCysGlyGlnIleIleSerGlnAspGlnAlaAspArgArgGlyLysValTyrA 659
   |||||
DB 5619 GAGATATATACGAGACACTGATCTCATGTAAGGATTAATGAGCGTGGAGAAATAGAG 5678
   |||||
QY 659 sPlyTyrMetCysSerPheLeuPheLeuAsn-----AsnAspPhe 673
   |||||
DB 5679 ATCGATTTGTTCTTCACTGCTTACCTTGAATGATCAGTAATTCAAGATTAATTT 5738
   |||||
QY 674 ValVal----- 675
DB 5739 GAAGTAAGCTTTTATATCATTCGCGGTTACACATCTATTCGAAATGAAATTAATTAT 5798
   |||||
QY 676 -----AspAlaThrArgLysGlnLysLysIleArgPheAlaAsnHisSer 690
   |||||
DB 5799 TTACAGCTCGAAATCGATGCTCGCCGTAAAGAAACGAGTTCAAAATTTCTCATATCA 5858
   |||||
QY 691 ValAsnProAsnCysTyrAlaLysVal----- 699
   |||||
DB 5859 GGAAGACCTTAATCTGCTACGCCAAGST-ACTAAGCCGTATATCTTATCTGAACAATAC 5917
   |||||
QY 699 ----- 699
DB 5918 TAACTATTAACAACAAATACTATATGTTTCTTATGTTAATCGTATCACTT 5977
   |||||
QY 699 ----- 699
DB 5978 TACTGCTGTTGATGTTGTTTCATATGAGATATTCAGAAACTCAAACTCATTTTAA 6037
   |||||
QY 699 ----- 699
DB 6038 ATGATTTTCTGTCGAGAAATTTAGCTTACGAAATTTATGCTTGTGTGAGCTTG 6097
   |||||
QY 700 MetMetValaAngLysAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGly 719
   |||||
DB 6098 ATGATTTGAGAGAGGATCAAGATGCTGATTTGCGAGAGAGCAATGAGAAAGST 6157
   |||||
QY 720 GlnGlnLeuPhePheAspTyrArgTyrSerGlnIleAspAla 733
   |||||
DB 6158 GAGAGCTTTTCTTCACTGCTATGCTATGAGCAACCAAGTCG 6199
   |||||

RESULT 14
US-09-812-283-6
Sequence 6, Application US/09812283
Patent No. 6828477
GENERAL INFORMATION:
APPLICANT: Fischer, Robert L.
APPLICANT: Ohad, Nir
APPLICANT: Kiyosue, Tomohiro
APPLICANT: Yadegari, Ramtin
APPLICANT: Margosian, Linda
APPLICANT: Harada, John
APPLICANT: Goldberg, Robert B.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
TITLE OF INVENTION: Development in Plants
FILE REFERENCE: 023070-086120US
CURRENT APPLICATION NUMBER: US/09/812, 283
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/177,249
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: US 09/071, 838
PRIOR FILING DATE: 1998-05-01
NUMBER OF SEQ ID NOS: 324
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 7015
TYPE: DNA

```

```
ORGANISM: Arabidopsis sp.
FEATURES:
  NAME/KEY: CDS
  LOCATION: (1)..(7014)
  OTHER INFORMATION: fertilization-independent endosperm 1 (FIR1)
  OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1
  NAME/KEY: CDS
  LOCATION: (2)..(7015)
  OTHER INFORMATION: fertilization-independent endosperm 1 (FIR1)
  OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2
  NAME/KEY: CDS
  LOCATION: (3)..(7013)
  OTHER INFORMATION: fertilization-independent endosperm 1 (FIR1)
  OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3
US-09-812-283-6

Alignment Scores:
Pred. No.: 5,12e-31      Length: 7015
Score: 436.00          Matches: 122
Percent Similarity: 34.54      Conservative: 42
Best Local Similarity: 25.64    Mismatches: 78
Query Match: 10.84          Indels: 235
DB: 3                  Gaps: 11

US-10-773-302-2 (1-746) x US-09-812-283-6 (1-7015)
QY 491 LysIlyblybArygLybHlArlLeuTrpAlaAlaHlscYbArygLybIlEgIlLeuLybLyb 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 4782 CGCAAAATGAGAGCTCCGAAATATATGCTGTTATCCGCCCTTTAAAGAAACAACT 4841
QY 511 AspGlySerSerAsnHlValIlyrAntYrgInProCyAspHlspRoArgGlnProCy 530
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 4842 AGTGAGAGAGCTAAGTTTATATAGCACTACACACATCACT---TGCAAGCAAAATGT 4898
QY 531 AspSerSerCyProCyValIlleAlaGlnAsnPhCyGlnLybPhCy 547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 4899 GAGACGACATGCCCTTGTTTAACTCAGCAAAATTCCTCGAGAAATATTGGGGTATGTC 4958
QY 547 ----- 547
DB 4959 ATTCAATTTTCTTACGCCGAGATCCATGATTTAATTGAAACATGATTTGATTT 5018
QY 548 -----Gln-CySerSerGluCyGlnAsnArpHeProGlyCybAryCybLyb---Al 564
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5019 TTTGTTCAAGTCTCAAAAGATTCGACATGCTTTGGAGAGATGTAATTGTGCAATTGG 5078
QY 564 aglncYbAsnHlLybGlnCybProCybTyrlEulAlaValaGlnGluCybAspProAspLe 584
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5079 CCAATGCACAAAATCGACAAATGCTCTGTTTGTCTGTAATCGTAATGCGATCCAGATCT 5138
QY 584 uCybLeuHrCybGlyAlaAlaAspHlETrpAsp----- 595
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5139 TTGTGGAGTTTCTCTTAAAGTAAACATTTCACTTCAATATCTTTTATACAAATTTCTA 5198
QY 596 -----SerLybAsn----- 598
DB 5199 TAATCAAAAGTATTCAAACAAAGCTTATAAATACTTATATATAGCTGTGAGAGA 5258
QY 599 -----ValSerCybLybAsn-CySer----- 605
DB 5259 TGGCACTTGTGTGAGACACAGTGCMAATTCMAATGCAAGACATGCAATTCCTCTTCA 5318
QY 606 -----IlEgIlArGlySerLybLybHl8----- 613
DB 5319 AACCAATTAAAGGTATACAGTCAAAATCCGTACGAAATTTAAACTATTTATACGA 5378
QY 614 -----LeuLeuEulAlaProSerAspVala 622
DB 5379 AAGACATTTAATCATATTTCCGATTTTATCTAGATTTCTCATGGAAGTCTGATGTTTC 5438
QY 622 lAgLyTyrgLybIlyleHeille----- 628
DB 5439 ATGATGGGGTGCATTTTATCATGGGTAAAGCAATCATGTAATATTAAGAAATAGTTTAATG 5498
```

```
QY 628 ----- 628
DB 5499 TTATTGGCATTCATAACATTTTATTTTATATATGATTTTATCTTACATTA 5558
QY 629 -----LysAspProValGlnLybAsnGlnHeille 639
DB 5559 AATATATTTGTGTATATGTTTGAACCGGTACAGACTCTCTTAAAGAAATGATATCTCG 5618
QY 639 ergIuTyCybGlyGlnIlleIleSerGlnAspGluAlaAspArgArgGlyLybValTyra 659
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5619 GAGAAATTTACTGAGAACTGATCTCATATGATAGAGCTTAATGACGTGGAGATAGAG 5678
QY 659 sPlyTyrmCybSerPheLeuPheAsnEulAsn-----AsnAspPhe 673
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5679 ATGGATTTGGTTCTTCTGACTCTTTACCTTGAATGATCAGTTAATTTCAAGAAATATTTT 5738
QY 674 ValVal----- 675
DB 5739 GAAGTAAAGTTTATATATTCGCGGTATACATCTATTCGAATCAAAAGTAACTTATTT 5798
QY 676 -----AspAlaThrArgLybGlybAsnLybIleArgPheAlaAsnHlSer 690
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5799 TTACAGCTCGAAATCGATGCTCGCGCTTAAGAAACAGATTCAATTTCTCAATCACTCA 5858
QY 691 ValAsnProAsnCybTyraAlaLybVal----- 699
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5859 GCAAGACTTAACTGCTAGCGCAAGT-ACTAAGCCGTTATCTTATCTTGAACAAATAC 5917
QY 699 ----- 699
DB 5918 TAACTATTAACAAACAAATATCTTATGTTAGTTTCTTTAAGTAAATCGTATCAACTT 5977
QY 699 ----- 699
DB 5978 TACTGTCGTGATTTGTTTCATATTTGAAGATATTCAGAAACTCAAACTCATTTTAA 6037
QY 699 ----- 699
DB 6038 ATGATTTTCTGTCGAGAAATTTAGTTACGAAATTTATGTTTCTGTGCAGTTG 6097
QY 700 MetHeValaAsnGlybAsnHlArlGlyIlyleHeAlaLybAryAlaIlEgIlThrGly 719
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 6098 ATGATTTGAGAGAGATCAGAGATTTGCTATTTTGGGAGAGACATGAAAGAGT 6157
QY 720 GluGlnLeuPhePheAspTyraTyrlSerGlnAlaAspAla 733
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 6158 GAGAGCTTTCTTGCATCTGCTATGACACAGAACTGG 6199

RESULT 15
US-09-071-838A-6
Sequence 6, Application US/09071838A
Patent No. 7029917
GENERAL INFORMATION:
APPLICANT: Fischer, Robert L.
            Onad, Nit
            Kiyosue, Tomohiro
            Yadegari, Ramin
            Margosian, Linda
            Harada, John
            Goldberg, Robert B.
TITLE OF INVENTION: Nucleic Acids That Control Seed and
                    Fruit Development in Plants
NUMBER OF SEQUENCES: 324
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
```



Qy 720 GluGluLeuPhePheAspTyrArgTyrSerGlnAlaAspAla 733  
|||  
Db 6158 GAGGAGCTTTCTTGGACTACTGCTATGACCGAACAATGGG 6199  
|||

Search completed: August 13, 2006, 17:30:09  
Job time : 396.311 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 13, 2006, 16:46:46 ; Search time 2194.12 Seconds  
(without alignments)  
6266.694 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGCTGKSEKGPVCMKRVK.....RYSQADALKYIGIEHREMRIP 746

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frames+ p2n.model -DEV=xlh  
-O=/abs/ABSSWB.spool/US10773302/runat.11082006.140239.16873/app\_query.fasta\_1  
-DB=Published Applications NA.Main -QFMT=fastap -SUPP=p2n.rnpbm  
-MINMATCH=0.1 -LOOFCR=0 -LOOPEXT=0 -UNITS=bits -STAR=1 -END=1  
-MATRIX=DIOSum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext  
-HEARSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs807  
-USER=US10773302 @CGN 1.1.2675 @runat.11082006.140239.16873 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THRPADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.Main:

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq:\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq:\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11E\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4051	100.0	2600	3	US-09-954-456-1157 Sequence 1157, Ap

2	4051	100.0	2600	10	US-10-843-641A-4184 Sequence 4184, Ap
3	4051	100.0	2600	10	US-10-773-302-1 Sequence 1, Appli
4	4051	100.0	2711	9	US-10-723-860-5458 Sequence 5458, Ap
5	4038	99.7	2558	6	US-10-153-668-231 Sequence 231, App
6	4038	99.7	2576	7	US-10-172-118-1043 Sequence 1043, Ap
7	4038	99.7	2576	7	US-10-295-027-709 Sequence 709, App
8	4038	99.7	2576	8	US-10-342-887-1043 Sequence 1043, App
9	4038	99.7	2576	10	US-10-848-755A-129 Sequence 129, App
10	4038	99.7	2576	10	US-10-504-173-78 Sequence 78, Appli
11	4038	99.7	2576	12	US-10-960-814-396 Sequence 396, App
12	4024	99.3	2512	7	US-10-210-120-95 Sequence 95, Appli
13	4024	99.3	2512	10	US-10-909-035-95 Sequence 95, Appli
14	3812.5	94.1	2476	7	US-10-104-047-1192 Sequence 1192, Ap
15	3812.5	94.1	2476	16	US-11-072-512-1192 Sequence 1192, Ap
16	3106	76.7	1960	6	US-10-153-668-333 Sequence 333, App
17	2606.5	64.3	1485	10	US-10-764-420-1602 Sequence 1602, Ap
18	2601.5	64.2	2594	10	US-10-504-173-77 Sequence 77, Appli
19	2429.5	60.0	2522	7	US-10-115-482-55 Sequence 55, Appli
20	2156.5	53.2	2448	13	US-11-097-143-14072 Sequence 14072, A
21	1932.5	47.7	5289	13	US-11-097-143-14071 Sequence 14071, A
22	1208.5	29.8	807	8	US-10-424-559-76875 Sequence 76875, A
23	1138	28.1	714	9	US-10-333-872A-67 Sequence 67, Appli
24	1056	26.1	555	3	US-09-796-692-5700 Sequence 5700, Ap
25	1056	26.1	555	6	US-10-040-862-5700 Sequence 5700, Ap
26	1056	26.1	555	7	US-10-057-475B-5700 Sequence 5700, Ap
27	1056	26.1	555	7	US-10-154-884B-5700 Sequence 5700, Ap
28	1056	26.1	555	9	US-10-764-324-5700 Sequence 5700, Ap
29	921	22.7	548	6	US-09-796-692-5422 Sequence 5422, Ap
30	921	22.7	548	6	US-10-040-862-5422 Sequence 5422, Ap
31	921	22.7	548	7	US-10-057-475B-5422 Sequence 5422, Ap
32	921	22.7	548	7	US-10-154-884B-5422 Sequence 5422, Ap
33	921	22.7	548	9	US-10-764-324-5422 Sequence 5422, Ap
34	835	20.6	3162	9	US-10-733-930-2112 Sequence 2112, Ap
35	828.5	20.5	2857	8	US-10-425-114-12944 Sequence 12944, A
36	813.5	20.1	3030	3	US-09-906-453-3 Sequence 3, Appli
37	813.5	20.1	3030	16	US-11-230-145-3 Sequence 3, Appli
38	809.5	20.0	3061	7	US-10-310-154-231 Sequence 231, Appli
39	808	19.9	2954	8	US-10-424-559-31171 Sequence 31171, A
40	797	19.7	3173	9	US-10-733-930-1796 Sequence 1796, Ap
41	797	19.7	3193	9	US-10-425-115-96409 Sequence 96409, A
42	790.5	19.5	3207	8	US-10-437-963-3210 Sequence 3210, Ap
43	784	19.4	3074	8	US-10-437-963-43565 Sequence 43565, A
44	783	19.3	3180	3	US-09-906-453-1 Sequence 1, Appli
45	783	19.3	3180	16	US-11-230-145-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-09-954-456-1157  
; Sequence 1157, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cai  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954, 456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233, 617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234, 052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234, 923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 711  
; PRIOR FILING DATE: 2000-09-27

```
; PRIOR APPLICATION NUMBER: US/60/235, 720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1157
; LENGTH: 2600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1157

Alignment Scores:
Pred. No.: 0 Length: 2600
Score: 4051.00 Matches: 746
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatch: 0
Query Match: 100.0% Indels: 0
DB: Gaps: 0

US-10-773-302-2 (1-746) x US-09-954-456-1157 (1-2600)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20
DB 90 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCAAGTTGTGGCCGAAACGGTAA 149
QY 21 SerGluTrpMetLysArgLysGlnLysLysArgPheArgAlaAspGluValLysSer 40
DB 150 TCAAGTACATGCGACCTGACAGACCTCAAGAGTTCAACCAAGCTGATGAAGTAAAGT 209
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60
DB 210 ATGTTTAGTCCAAATCGTCAGAAATTTTGGAAAGACGGAATCTTAAACCAAGAAATG 269
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80
DB 270 AAACCGCGAAGATACAGCTGTCGACATCTGACTTCGTGAGCTCATTCGCGGAGCT 329
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100
DB 330 AGGAGTGTTCGGTGAACAGTGAAGCTTGGAATTTTCCAAACAAGTCATCCATTAAAGCT 389
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnAsnPhe 120
DB 390 CTGAATGAGTTGCTTCAGTACCATTAATGTAATCTTGTCTCCCTACACAGAAATTTT 449
QY 121 MetValGluAspGluThrValLeuHisAsnIleProLysMetGlyAspGluValLeuAsp 140
DB 450 ATGGCGAAGATGAAACTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTAT 509
QY 141 GlnAspGlyThrPheIleGlnGluLeuLysAsnLysAsnLysAsnLysValHisGlyAsp 160
DB 510 CAGAGTGTACTTTCATTGAAAGACTTAATTAATTAATTAATTAATTAATTAATTAAT 569
QY 510 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGGTGAATGCCCTTGCTGCA 629
QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200
DB 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220
DB 690 AAAGTCTGGAGATCACCGAGATATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 749
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240
DB 750 GATAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGCAGCAGGAAGAACTA 809
QY 241 LysGluLysTrpLysGluLeuThrGluGlnLeuProGlyAlaLeuProGluCys 260
```

```
DB 810 AAGGAAAAATTAAGAAGCTCACCAACAGAGCTCCACAGCGCACTTCTCTGATGT 869
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnInsLeuHisSer 280
DB 870 ACCCCCAACATTAATGAGCAACAAATCGTAAATCTGTTCAAGAGAGCAAACTTACCTCC 929
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisProPheHis 300
DB 930 TTTCAATGCGCTTTTCTTATAGCCATGTTTAAATATATACGTCTTCAATCTTTTCAT 989
QY 301 AlaThrProAsnThrTrpLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320
DB 990 GCACACCCCAACATTAATTAAGCGAAGAACACAGAAACAGCTTACCAACAACTTGT 1049
QY 321 GlyProGlnCysTrpGlnHisLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340
DB 1050 GGACCAAGTGTATCAAGCATTTGGAGGAGCAAGAGATTGTGCTGCTTCAACCGCT 1109
QY 341 GluArgIleLysThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360
DB 1110 GAGCGATTAAGACCCCAACAAACGTCCAGAGGCGCAGAAAGAGACGGCTTCCAAT 1169
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380
DB 1170 AACGTTACAGGCCCAAGCAACCCCAACATTAATGTCTGAATCAAGATACAGACGT 1229
QY 381 AspArgLysAlaGlyThrGluThrGlyGlyLysAsnAsnAspLysGluGluGluLys 400
DB 1230 GATAGGAGACAGGAGCTGAAGACGGGGAGAGAAACAATGATTAAGAAAGAAAGAGAG 1289
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420
DB 1290 AAAGATGAACCTTCGAGACTCTCTGAAGCAAAATCTCGGTGTCAACCAATTAAGATG 1349
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440
DB 1350 AAGCCAAATTTGAACCTCTGAGAAATGTGAAGTGGAGTGTGCTGAAGCTCAATTTT 1409
QY 441 ArgValLeuIleGlyThrTrpTrpAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460
DB 1410 AGAGTCTCATTTGCACTTACTATGACAATTTGTGTGCTTACATTTGTTAATTTGGAGCC 1469
QY 461 LysThrCysArgGlnValIleTrpGluPheArgValIleGluSerSerIleIleAlaProAla 480
DB 1470 AAAACATGTAGACAGGGTGATGAGTGAAGTCAAAACCTAGACATCATAGCTCCAGCT 1529
QY 481 ProIleGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500
DB 1530 CCGCTGAGATGTGATTAATCTCTCCAGAAAGAAAGAGAGAAACACCGGTTGTGGGCT 1589
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspLysSerSerAsnHisValTyrAsnTrp 520
DB 1590 GCACACTGCAAGAAAGATACAGCTGAAGAAAGAGCGGCTCTCTAACATGTTTACAACTAT 1649
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB 1650 CAACCTGTGATCATTCACAGCAGCTTGAGACAGTTGCGCTTGTGTGATGACAA 1709
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGlyCys 560
DB 1710 AATTTTGTGAAAGATTTTGTGCAATGTATGATTCAGAGTGTCAAAACCGCTTCCGGATGC 1769
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValArgLysCys 580
DB 1770 CGCTGCAAGACAGTCAACACCAAGCAGAGCCGGGTACTACCTGGCTGTCCGAGAGTGT 1829
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisTrpAspSerLysAsnValSer 600
DB 1830 GACCTGACCTGTGTCTTACTTGTGAGCGGTGACCAATTTGGGACATTAAGTAAGTGTCC 1889
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAsp 620
```

Db 1890 TCAGAAAGCTGAGTATTCAGCGGCTCCAAAAGCATCTATTCGTGCGACCATCTGAC 1949  
Qy 621 VALAAGIYTRGIIYTLepheileysAspProValGlnIlyAsnGlnIlyleSergIu 640  
Db 1950 GTGGCAGGCTGGGGATTTTATCAAGATCCGTGCAAAAATCAATTCATCTCAGAA 2009  
Qy 641 TTYCSGIIYGLIILEISergIuAspGluIlyAspArgGIIYIlyValTYrAspIlyS 660  
Db 2010 TACTGTGAGAGATTTTCTCAAGATGAGCTGACAGAGAGGAAATGTGATGATAA 2069  
Qy 661 TTYmetCySerPheLeuPheAsnIleuAsnAspPheValIlyAspAlathrArgIlyS 680  
Db 2070 TACATGTGACGCTTCTGTTCACCTTGAACATGATTTGTGTGATGATCAACCGCAG 2129  
Qy 681 GIIYAsnIlySIIleArgPheIlyAsnIlySserValAsnProAsnCyTYrAlalyValIlyMet 700  
Db 2130 GGTAACTAAATTCGTTTGCAATCATCTGGTAAATCCAACTGCTATCCAAAGTTAAG 2189  
Qy 701 MetValAsnGIIYAspIlySArgIlyleGIIYIlyPheAlalySArgAlalyIleGIIYThrGIIYIly 720  
Db 2190 ATGCTTAACGTGATCAACGATAGTATTTTTCAGAGAGAGCCATCCAGCTGCGAA 2249  
Qy 721 GIIleuPhePheAspTYrArgTYrSergIuIlyAspAlalySArgIlyIlyleGIIY 740  
Db 2250 GAGCTGTTTGTGATTAAGATACAGCCAGGCTGATGCCCTGAAGTATGCGCATCGAA 2309  
Qy 741 ArgGIIYMetGIIYIlyPro 746  
Db 2310 AGAGAAATGGAAATCCCT 2327

RESULT 2  
US-10-843-641A-4184  
; Sequence 4184, Application US/10843641A  
; Publication No. US2005064454A1  
; GENERAL INFORMATION:  
; APPLICANT: Avalon Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; TITLE OF INVENTION: Signature Gene Sets  
; FILE REFERENCE: 689290-189  
; CURRENT APPLICATION NUMBER: US/10/843,641A  
; CURRENT FILING DATE: 2004-05-12  
; PRIOR APPLICATION NUMBER: US/09/873,367  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US/09/954,531  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,436  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,832  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/964,824  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/09/967,768  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US/09/968,007  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,347  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,708  
; PRIOR FILING DATE: 2001-10-03  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 8447  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4184  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-843-641A-4184

Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 4051.00 Matches: 746

Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 10 Gaps: 0

US-10-773-302-2 (1-746) x US-10-843-641A-4184 (1-2600)

Qy 1 MetGIIYIlyThrGIIYIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlyS 20  
Db 90 ATGGCCAGACTGGGAGAAATCTGAGAAAGGAGCAAGTTGTGGCGAGAGCTGTAATAA 149  
Qy 21 SergIuTYrMetArgLeuArgGIIYIlySArgPheArgArgAlalySergIuIlySergIu 40  
Db 150 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTCAAGACAGCTGATGAAGTAAAGGT 209  
Qy 41 MetPheSerSerAsnArgGIIYIlySergIuIlySergIuIlySergIuIlySergIuIlyS 60  
Db 210 ATGTTAGTCCAACTGTCAGAAATTTTGGAAAGAGAAATCTTAAACCAAGAAATGG 269  
Qy 61 IlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIu 80  
Db 270 AAACGCGAAGAGATACAGCTGTCAGATCTGAGCTCATGCGCGAGACT 329  
Qy 81 ArgGIIYCySerValTYrSerAspLeuAspPheProThrGIIYIlyProleuIlyThr 100  
Db 330 AGGAGTGTTCGTGACAGTACAGTATTTTCCAAACAGATCATCCATTAAGACT 389  
Qy 101 LeuAsnAlalyAlalySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlyS 120  
Db 390 CTGAATGAGATGCTTCAGTACCTCAATATGATTCGTGCTCCCTTACAGCAGAAATTTT 449  
Qy 121 MetValIlyAspGIIYIlyThrValIlyIlySergIuIlySergIuIlySergIuIlySergIu 140  
Db 450 ATGGTGAAGAGTAACTGTTTACATTAACATCTTATATGAGAGAGAGTTTATGAT 509  
Qy 141 GIIAspGIIYThrPheIlyGIIYIlySergIuIlySergIuIlySergIuIlySergIuIlyS 160  
Db 510 CAGATGTACTTTCATTTGAGAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 569  
Qy 161 ArgGIIYCySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlyS 180  
Db 570 AGAGATGTGGGTTTATGATGAATGAAATTTTGGAGTTGGTGAATGCCCTTGSTCAA 629  
Qy 181 TYrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
Db 630 TATAATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
Qy 201 IlySAspLeuGIIYAspIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlyS 220  
Db 690 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
Qy 221 AspIlySIIlePheGIIYIlySergIuIlySergIuIlySergIuIlySergIuIlySergIu 240  
Db 750 GATAAATTTTGAAGCATTTCTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 809  
Qy 241 IlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIu 260  
Db 810 AAGGAAATTAATAAGAACTACCGAAGAGAGCTCCAGGCGCACTTCTCTGAAATGT 869  
Qy 261 ThrProAsnIlyAspGIIYProAsnAlalySergIuIlySergIuIlySergIuIlySergIu 280  
Db 870 ACCCCCAACATGATGAGCAACCAATGCTTAATCTGTTCAAGAGAGCAAGCTTACACTCC 929  
Qy 281 PheIlySIIlePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 300  
Db 930 TTTCAATGAGCTTTCTGAGGCGATGTTTAAATATGACTGCTCTCTCATCTTTTAT 989  
Qy 301 AlathrProAsnIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIu 320  
Db 990 GCAACACCAACACTTATAGCGAGAGAAACAGAAACAGCTCTAGAACAAACCTTGT 1049  
Qy 321 GIIYProGIIYIlySergIuIlySergIuIlySergIuIlySergIuIlySergIuIlySergIu 340

Db 1050 GAACACAGTGTACAGACATTTGGAGGAGCAAAAGAGTTGCTGCTCTCAACGGCT 1109  
Qy 341 GUAATGILEYSTRPProProLyArpProGlyArgArGArGlyArgLeuProAn 360  
Db 1110 GAGCGGATTAAGACCCCAAAACGTCCAGAGAGCCGCAAGAGAGCGCTTCCCAAT 1169  
Qy 361 AmsSerArpProSerThProThrLeaAnValLeuGlySerLyArpThArpSer 380  
Db 1170 AACGTAGCAGAGCCAGCAACCCCACTTAATGTGCTGAATCAAGATACAGACAGT 1229  
Qy 381 AspArpGluAglYThrGluThrGlyGlyGluAnAAspLyGluGluGlyLys 400  
Db 1230 GATAGGAGAGCGAGGATGAAACGGGGAGAGAAACAATGATTAAGAAAGAGAGAG 1289  
Qy 401 LyAAspGluThrSerSerSerSerGlyAaAmsSerArpCySerGlnThProIleYMet 420  
Db 1290 AAAATGAAACTTGAGAGCTCTCTGAACAATTCGTGGTTCAAACGCAATAAAGATG 1349  
Qy 421 LysProAenIleGluProGlyAnValGluTPSerGlyAglValAspMetPhe 440  
Db 1350 AAGCAATATATTGAACCTCTGAGAAATGTGAGTGGTGGTGAAGCCTCAATGTTT 1409  
Qy 441 ArgValLeuIleGlyThrYThrYArpAAspPheCyAaIleAArgLeuIleGlyThr 460  
Db 1410 AGAGTCTCATTTGGCATTAATGACAAATTCGTGCATTTGCTAGGTAAATGGAGCC 1469  
Qy 461 LysThrCyAAspGlnValYArgLupPheArgValLysGlySerSerIleIleAProA 480  
Db 1470 AAAACATGTACAGAGGTATAGAGTTAGAGTCAAAAGAAATCAGATCATAGCTCAAGCT 1529  
Qy 481 ProAaglUAspValAspThProProArgLyLyLyLyValGlyLysArgLeuTPAa 500  
Db 1530 CCCCCTGAGATGTGATATCTCTCAAGAAAGAGAGAAACCCGGTGTGGGCT 1589  
Qy 501 AlaHisCyAAspGlyIleGluLeuLysYAspGlySerSerAAspHisValYArgPhe 520  
Db 1590 GCAACTGCAAGAAATACAGTGAAGAAAGAGCGCTCTCTTACCAATGTTTCAACTAT 1649  
Qy 521 GlnProCyAAspHisProArgGlnProCyAAspSerSerCyProCyValIleAArgL 540  
Db 1650 CAACCTGTGATCATCCAGGAGAGCTTGTGACAGTGTGTCCTGTGTATAGCAACA 1709  
Qy 541 AAspPheCyGlyLupPheCyGlnCySerSerGlyLupGlnAAspPheProGlyCyS 560  
Db 1710 AATTTTGTGAAAGTTTGTCAATGATGTCAAGTGTCAAAACCCGTTCCGGGAGGC 1769  
Qy 561 ArgCyLyValAglInCyAAspThThrArgGlnCyProCyBUTYrLeuAValArgGluCyS 580  
Db 1770 CGCTGCAAGCAAGTGAACCAAGCAAGTGCCTGCTACCTGCTGCTGCGAGAGTGT 1829  
Qy 581 AspProAAspLeuCyLeuThrCyGlyValAaAAspHisArTPAspSerLyAAspValSer 600  
Db 1830 GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGGACAGTAAATGTGTCC 1889  
Qy 601 CyLyValAAspCySerIleGlnArgGlySerLyLyLysLeuLeuAAspProSerArp 620  
Db 1890 TGCAAGAACTGCAAGTATCAGCGGGCTCCAAAAGCAATCTATGCTGCGCAACTCAGC 1949  
Qy 621 ValAaglYTPGlyIlePheIleYAspAAspProValGlnLysAAspGluPheIleSerGlu 640  
Db 1950 GTGGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAAATGAAATTCATCTCAANA 2009  
Qy 641 TYrCyGlyGlyLulIleSerGlnAAspGluAaAspArGArGlyLysValTYrAspLyS 660  
Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAAAGAGGGAAGGTATATATAA 2069  
Qy 661 TYrMetCySerPheLeuPheAAspLeuAAspAAspPheValValAspAlaThrArgLyS 680  
Db 2070 TACATGTGACAGCTTCTGTCAACTTGAAACATGATGTTTTGTGTGAGATGCAACCCGCAAG 2129  
Qy 681 GlyAAspLySIIeArgPheAlaAspHisSerValAAspProAAspCyTYrAlaLysValMet 700  
Db 2130 GGTAACAAAATTCGTTTGCAATATTCGGTAATTCAAACATGCTAATGCAAAAGTTATG 2189

Qy 701 MetValAAspGlyAAspHisArgGlyIleGlyIlePheAlaLysArgAlaIleGlnThrGlyL 720  
Db 2190 ATGGTTAAACGCTGATACAGAGATAGATATTTTGTCCAAAGAGAGCTCAAGACTGGCGAA 2249  
Qy 721 GluLeuPhePheAspTYrArgTYrSerGlnAaAspAlaLeuLYrTYrValGlyIleGly 740  
Db 2250 GAGCTGTTTTTGTATTAACAGTACAGCCAGGCTGATCCCTGAAGTATGTGCGCATGGA 2309  
Qy 741 ArgGluMetGluIlePro 746  
Db 2310 AGAGAAATGGAAATCCCT 2327  
RESULT 3  
US-10-773-302-1  
; Sequence 1, Application US/10773302  
; Publication No. US2005008980A1  
; GENERAL INFORMATION:  
; APPLICANT: Jenuwein, Thomas  
; APPLICANT: Laible, Gocz  
; APPLICANT: O'Carroll, Donal  
; APPLICANT: Eisenhaber, Frank  
; APPLICANT: Rea, Stephen  
; TITLE OR INVENTION: Chromatin-Regulator Genes  
; FILE REFERENCE: 0652.167001  
; CURRENT APPLICATION NUMBER: US/10/773.302  
; PRIOR FILING DATE: 2004-02-09  
; PRIOR APPLICATION NUMBER: US/09/589.892  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US 08/945.988  
; PRIOR FILING DATE: 1997-11-10  
; PRIOR APPLICATION NUMBER: PCT/EP96/01818  
; PRIOR FILING DATE: 1996-05-02  
; PRIOR APPLICATION NUMBER: DE 195 16 776.7  
; PRIOR FILING DATE: 1995-05-10  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURES:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)..(89)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (90)..(2330)  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: (2331)..(2600)  
US-10-773-302-1  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0  
US-10-773-302-2 (1-746) x US-10-773-302-1 (1-2600)  
Qy 1 MetGlyGlnThrGlyLyLySerGlyLupGlyProValCyBTPArGlyAAspValLys 20  
Db 90 ATGGGCAAGACTGGGAAGAAATCTGAAGAAGGACCACTTGTGTGCGGAAGCGGTAAAA 149  
Qy 21 SerGluTYrMetAArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
Db 150 TCAAGATGACATGCACTGAGCAAGCTCAAGAGGTTTCAGACAGAGCTGATGAAGTAAAGGT 209  
Qy 41 MetPheSerAAspAAspGlnLYrIleGluGluArgThrGluIleLeuAAspGlnGluTP 60  
Db 210 ATGTTTACTTCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAACCAAGAAATGG 269



APPLICANT: Aziz, Natasha  
APPLICANT: Ginsburg, Wendy M.  
APPLICANT: Zlotnick, Albert  
TITLE OR INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
TITLE OR INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
FILE REFERENCE: 05882, 0193, NEUS01  
CURRENT APPLICATION NUMBER: US/10/723, 860  
CURRENT FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429, 739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 5458  
LENGTH: 2711  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-773-860-5458

Alignment Scores:  
Pred. No.: 0 Length: 2711  
Score: 4051.00 Matches: 746  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0

US-10-773-302-2 (1-746) x US-10-723-860-5458 (1-2711)

QY 1 MetGlyGlnThrGlyLysLeuSerGlyLysProValCysTrpArgLysArgValLys 20  
DB 167 ATGGCCGACCTGGGAAAGAAATCTGAGAGGAGCACTTGTGGCGAGCGGTAA 226  
QY 21 SerGlyTrpMetArgLeuArgGlnLeuLysArgPheArgAlaAspGlnValLysSer 40  
DB 227 TCAGAGTACATCCCACTGACGACGCTCAAGAGCTTCAACAGCTGAGAAAGTAAAGCT 286  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnLysTrp 60  
DB 287 ATGTTAGTCCAACTGTCAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 346  
QY 61 LysGlnArgArgGlnLeuProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACGCCAAGAGATACACCTGTCACATCTGACTTGTGAGCTCATGGCGGAGCT 406  
QY 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 407 AGGAGGTTCGGTACCAAGTACCTGATTTTCCAAACAGTACATCCATTAAAGACT 466  
QY 101 LeuAsnAlaValAlaSerValProIleMetCysTrpSerProLeuGlnAsnPhe 120  
DB 467 CTGAATGCAATGCTTCACTGATCCCAATATGATCTGCTCCCTTACAGACAAATTTT 526  
QY 121 MetValGlnAspGlnThrValLeuHisAsnIleProCysMetGlnArgGlnValLeuAsp 140  
DB 527 ATGGTGAAGAGAACTGTTTACATTAATCTTATATGAGGAGTGAAGTTTAAAT 586  
QY 141 GlnAspGlyThrPheIleGlnLysLeuLysAsnIleLysAsnIleLysValIleGlnLys 160  
DB 587 CAGAGTGTACTTCTTCACTTGAAGAACTAAATTAATTAATTAATTAATTAATTAATTAAT 646  
QY 161 ArgGlnCysGlyPheIleAsnAspGlnLysPheValGlnLeuValAsnAlaLeuGlnLys 180  
DB 647 AGAGAAATGCGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 706  
QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
DB 707 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
QY 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 767 AAAGATCTGGAGATCAACCGAGATGATTAAGAAACCGCCCACTTCGGAATTTCTTCTT 826  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLysLeu 240

DB 827 GATTAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGACACAGAAAGAACTA 886  
QY 241 LysGlnLysTrpLysGlnLeuThrGlnGlnIleLeuProGlnValIleLeuProGlnCys 260  
DB 887 AAGGAAATTAATTAAGAACTCAACGAAACAGACACTCCAGGCCCACTTCTCTGAATGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnLysLeuHisSer 280  
DB 947 ACCCCCAACATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGCAACACTTACCTCC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisProPheHis 300  
DB 1007 TTTCAATAGCTTTTGTAGTACATGTTTAAATATACCTGCTTCTTCAATCTTTTCAAT 1066  
QY 301 AlaThrProAsnThrTrpLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB 1067 GCACACCCCAACCTTAAAGCCGAAAGAACAGAAACAGCTCTAGACAAACAACTTGT 1126  
QY 321 GlyProGlnCysTrpGlnHisLeuGlnGlnValLysGlnPheAlaAlaAlaLeuThrAla 340  
DB 1127 GACACACAGTGTTAACAGCACTTGGAGGAGCAAGAGATTTGCTGCTCTCACCGCT 1186  
QY 341 GlnArgGlnLysThrProProLysArgProGlnGlnValArgArgArgGlnArgLeuProAsn 360  
DB 1187 GAGCGATTAAGAACCCCAACCAACGTCACAGAGCCGCGAGAGAGAGAGAGAGAGAGAG 1246  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnLysLeuAspThrAspSer 380  
DB 1247 AACAGTACAGAGCCCAACCAACCCCAACCAATTAATGCTGGAATCAAGATTAACAGACT 1306  
QY 381 AspArgGlnAlaGlnLysThrGlnLysGlnLysAsnAspLysGlnGlnLysGlnLys 400  
DB 1307 GATGAG 1366  
QY 401 LysAspGlnThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1367 AAAGATTAATCTGACCTCTCTGAGCAAAATTTCTGCTGCTCAACCAATTAATGATG 1426  
QY 421 LysProAsnIleGlnProProGlnAsnValGlnTrpSerGlnAlaGlnAlaSerMetPhe 440  
DB 1427 AAGCAATATTAAGTAACTCTCGAAGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1486  
QY 441 ArgValLeuIleGlnLysThrTrpAspAsnPheCysAlaAlaIleAlaGlnLysIleGln 460  
DB 1487 AGAGTCTCAATGGCAGCTTACATTAAGCAATTTCTGCTCAATGCTTGAATTAATGGAG 1546  
QY 461 LysThrCysArgGlnValIleTrpGlnPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB 1547 AAACATGATGACAGGTGATGAGTTTGAAGTCAAAAGAACTTACATCAATGCTCCAGCT 1606  
QY 481 ProAlaGlnAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
DB 1607 CCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1666  
QY 501 AlaHisCysArgGlnLysIleGlnLysLysLysLysLysLysLysLysLysLysLysLys 520  
DB 1667 GCACACTGCAAGAAAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1727 CAACCTGTGATCACTCAACGAGCACTTGTGACAGTTCTGTCCTTGTGTGATGACAA 1786  
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlnCys 560  
DB 1787 AATTTTGTGAAAGTTTGTCAATGTATGATTAAGTCAAGTCAAAACCGCTTCCGGAGTGC 1846  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValArgGlnCys 580  
DB 1847 CGCTGCAAGAGCAGTGTGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1906  
QY 581 AspProAspLeuCysLeuThrCysGlnValAlaAlaAspHisTrpAspSerLysAsnValSer 600







Db 358 CTGATGAGTGGCTTCACTGACCACTAATGATTTCTGGTCTCCCTCAGACGAAATTT 417  
 Qy 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
 Db 418 ATGGTGAAGAGAAAGTGTGTACATTAACATTCCTTATATGGAGATGCAATGATTTAGAT 477  
 Qy 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyValValHisGlyAsp 160  
 Db 478 CAGGATGTGACTTCACTGAGAACTAATAAATAATATGATGGAGAAAGTACACGGAGAT 537  
 Qy 161 ArgGluGlyPheIleAsnAspGluIlePheValGluLeuValAsnIleLeuGlyGln 180  
 Db 558 AGAGATGTGGCTTTATATAATGATGAATTTTGTGGAGTTGGTGAATCCCTTGGTCA 597  
 Qy 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluValGluGluGln 200  
 Db 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 Qy 201 LysAspLeuGluAspHisAspAspAspLysGlySerAspProGluGluPheProSer 220  
 Db 658 AAAGATGTGGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAATTTCTTCT 717  
 Qy 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
 Db 718 GATATAATTTTGGAGGCCATTTCTCATGTTTCCAGATTAAGGCAACGACAGAAACTA 777  
 Qy 241 LysGluLysTyrLysGluLeuThrGluGluGlnLeuProGlyValAlaLeuProGluGly 260  
 Db 778 AAGAAATAATTAAGAACTACCGAACAGCAGCTCCGAGCGCATCTTCTCTGATAGT 837  
 Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnAspGluGlnSerLeuHisSer 280  
 Db 838 ACCCCCAACATGATGAGAACCAATCTAATCTGTTCAGAGACCAAGACTTACATCC 897  
 Qy 281 PheHisThrLeuPheCysAspArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
 Db 898 TTTCTACGCTTTCTGTAGCGCATGTTTAATATGACGTCTTCTACATCTTTTCAAT 957  
 Qy 301 AlaThrProAsnThrTyrLysAspGlyLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 Db 958 GCAACACCAACATTAATTAAGCGAAGAAACAGAAACAGCTCTGACACCAACCTTGT 1017  
 Qy 321 GlyProGlnCysTyrGlnHisIleLeuGluGlyValAlaLysGluPheAlaAlaIleThrAla 340  
 Db 1018 GACACACAGTGTACAGCATTTGAGAGGAGCAAGAGGTTGCTGCTCTTCAACCT 1077  
 Qy 341 GluArgIleLysThrProProLysAspProGlyLysArgArgGlyValArgLeuProAsn 360  
 Db 1078 GAGCGATTAAGACCCCAACCAACACCTCCAGAGGCCCGACAGAGAGACGCTTCCCAAT 1137  
 Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 Db 1138 AACAGTACAGAGCCAGACACCCCAACATTAATGCTGGAATCAAGAGATACACAGCT 1197  
 Qy 381 AspArgGluAlaGlyThrGluThrGlyGlyLysAsnAspLysGluGluGluGlyLys 400  
 Db 1198 GATAGGGAAGCAGGACGTAAGAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAG 1257  
 Qy 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 Db 1258 AAAGATGAACCTTCAGACTCTCTAGAGCAAAATTCCTGGTGTCAACACCAATTAAGATG 1317  
 Qy 421 LysProAsnIleGluProGluGluAsnValGluTyrSerGlyValGluAlaSerMetPhe 440  
 Db 1318 AAGCAAAATATTAACCTCTGAGAAATGAGAGTGAAGTGGTGAAGCTCAATGTTT 1377  
 Qy 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1378 AGAGTCCCTCATTTGGCATTAATACAAATTCCTGTGCACTTGCAGGTATTAATGGGACC 1437  
 Qy 461 LysThrCysArgGluValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480

Db 1438 AAAACATAGACAGTGTATGATGATTAAGTCAAAAGATCTAGCATATGCTCAGCT 1497  
 Qy 481 ProIleGluAspValAspThrProProArgLysLysValGlyValHisArgLeuThrAla 500  
 Db 1498 CCGCTGAGATGAGTGAATCTTCCAGAGAAAAAGAGAGAAACACCGGTGTGGGCT 1557  
 Qy 501 AlaHisCysArgLysIleGluLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 Db 1558 GCACCTGCAAAAGATACAGCTGAATAAAGAGCGCTCTCTTAACATGTTTACAACTAT 1617  
 Qy 521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 1618 CAACCTGTGATCATCCACGGCAGCTGTGACAGTGTGTCCTGTTGTATAGACADA 1677  
 Qy 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGluCys 560  
 Db 1678 AATTTTGTGAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
 Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLysValAlaArgGluCys 580  
 Db 1738 CCGTGCAGACAGCTGACACCAACAGAGTCCCGCTCTACCTGCTGCGAGAGTGT 1797  
 Qy 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisIlePheAspSerLysAsnValSer 600  
 Db 1798 GACCTGACCTGTGCTTACTTGTGAGCCGCTGACATGGGACAGTAAATGTGTCC 1857  
 Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleLeuLeuAlaProSerAsp 620  
 Db 1858 TGCAAGAACTCAGATATCACGGGGCTCCAAAAGCATTTATGCTGGCACACTGAC 1917  
 Qy 621 ValAlaGlyTyrPylIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1918 GTGGAGGCTGGGGATTTTATCAAAAGTCTGTGCAAAAATGAATTCATCTCAGAA 1977  
 Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 Db 1978 TACGTGAGAGATTAATTTCAAGATGAACCTGACAGAAAGGGAAGTGTATGATAA 2037  
 Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
 Db 2038 TACATGTGCACCTTCTGTTCACTTGAACATGATTTTGTGGTGAAGCAACCGCAG 2097  
 Qy 681 GlyAsnLysIleAspPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
 Db 2098 GGTAAACAAATTCGTTTGCATCAATTCGTTAAATCCAACTGATACCAAGTTATG 2157  
 Qy 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
 Db 2158 ATGTTAAAGGTGATACAGATAGTATTTTCCAAAGAGAGCATCAGACTGGCGAA 2217  
 Qy 721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740  
 Db 2218 GAGCTGTTTGTGTTACAGATACAGCAGCGCTGATGCCGTAAGTATGTGGCATCGAA 2277  
 Qy 741 ArgGluMetGluLeuPro 746  
 Db 2278 AGAGAAATGAGAAATCCCT 2295

RESULT 7  
 US-10-295-027-709  
 ; Sequence 709, Application US/10295027  
 ; Publication No. US20030232350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Afar, Daniel  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsberg, Wendy M.  
 ; APPLICANT: Glab, Kurt C.  
 ; APPLICANT: Glyme, Richard  
 ; APPLICANT: Hevezl, Peter A.  
 ; APPLICANT: Mack, David H.  
 ; APPLICANT: Murray, Richard  
 ; APPLICANT: Watson, Susan R.  
 ; APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/0.0/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
Remaining Prior Application data removed - See file wrapper or PAM.  
NUMBER OF SEQ ID NOS: 1366  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 709  
LENGTH: 2576  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-295-027-709

Alignment Scores:  
Pred. No.: 0 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.74 Conservative: 0  
Best Local Similarity: 99.74 Mismatches: 2  
Query Match: 99.74 Indels: 0  
Gaps: 0

US-10-773-302-2 (1-746) x US-10-295-027-709 (1-2576)

1 MetGlyIntnGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
58 ATGGCCAGACTGGAGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAGCTTAA 117  
21 SerGlyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
118 TCAGAGTACATCGACTGACAGACAGCTCAAGAGTTCAAGACGCTGAGTAAAGTAAAG 177  
41 MetPheSerSerAsnArgGlnLysIleLeuGluArgGlnIleLeuAsnGlnGlnIle 60  
178 ATGTTAATGTTCCAAATCGTCAGAAATTTTGGAAAGACGAAATCTTAAACGAAGATG 237  
61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgLysThr 80  
238 AAACGCCAAGATACAGCTCTGACATCTGATCTCTGAGCTCATTTGGCGGAGACT 297  
81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
298 AGGAGGTGTTCCGTGACCGAGTGAATTTTCCAAACAGATCATCCCATTAAGACT 357  
101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnIleAsn 120  
358 CTGAATGACAGTTGCTTCAGTACCCATTAATGTAATCTTGGTCTCCCTACAGCAATTTT 417  
121 MetValGluAspGlyThrValLeuHisAsnIleProLysMetGlyAspGluValLeuAsp 140  
418 ATGGTGGAGATGAATCTTTTACATAATCTTCTATATGCGAGATGAATTTTAAAT 477  
141 GlnAspGlyThrPheIleGlnIleLeuLysAsnLysAspGlyLysValHisGlyAsp 160

478 CAGATGTACTCTTCATTTAGAAACTATATATATATATATGAGAAAGTACGGGAGT 537  
161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
538 AGAGATGTGGGTTTATTAATGATGAAATTTTGTGAGTTGGTGAATGGCTTGGTCAA 597  
181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluGluGluGluGln 200  
598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
201 LysAspLeuGluAsnHisArgAspAspGluSerArgProAspGlyLysPheProSer 220  
658 AAATATCTGAGAGATCCAGAGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 717  
221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
718 GATTAATTTTGGAGGCAATTTCTCTCAATGTTTCCAGATTAAGGSCACAGCAGAACTA 777  
241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGluValAlaLeuProGluCys 260  
778 AAGGAAAAATATAAGAACTCACCCGAAACAGAGCTCCAGGCGCATCTCTCGAAATGT 837  
261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnIleArgGluGlnIleSer 280  
838 ACCCCCAACATATGAGATGACCAATGCTTAATCTGTGAGAGAGCAAGGCTTACCTCC 897  
281 PheIleThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
898 TTTCAATGCGTTTCTGTAGGCGATGTTTAAATATATGATGATGATGATGATGATGAT 957  
301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
958 GCAACACCCAAACACTTATTAAGCCGAGAGAACACAGAACAGCTCTAAGCAACAACTTGT 1017  
321 GlyProGlnCysTyrGlnHisIleLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340  
1018 GAGCCACAGGTATCCAGCATTTTGAAGAGAGCAAGAGATTTGCTGCTCTCAACGCT 1077  
341 GluArgIleLysThrProProLysArgProGluGlyLysArgArgGlyLysLeuProAsn 360  
1078 GAGCGGATTAAGATCCCAACCAAAAGCTCCAGAGCCGCGAGAAAGAGCGGCTTCCCAAT 1137  
361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
1138 AACAGTACAGAGCCAGACCCCAACCATTAATGTGCTGAAATCAAGATACAGACAGT 1197  
381 AspArgGluAlaGlyThrGlnThrGlyGlyGluAsnAspLysGluGluGluGluGlu 400  
1198 GATAGGAG 1257  
401 LysAspGluThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
1258 AAAGATGAATCTTGAAGTCTCTGAGCAAAATCTCGGTGTCAACCCATTAAGATG 1317  
421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
1318 AAGCCAAATATTAAGCTCTCGAGAAATGTGAGAGAGAGTGTGTGTAAGCTCAATGTTT 1377  
441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
1378 AGAGTCTCATTTGGACTTACTATGACAAATTTCTGTGCTAGTTGATTAATGGAGACC 1437  
461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
1438 AAACATGTATACAGGTGTATAGATTAAGATCAAAAGATATACATCATGCTCCAGCT 1497  
481 ProAlaGluAspValAspThrProProArgLysLysLysValGlyHisArgLeuThrAla 500  
1498 CCGCTGAGATGTGAGATACCTCTCCAAAGGAAAAAGAGAGAAACACCGGTTTGGGCT 1557  
501 AlaHisCysArgLysIleGlnLeuLysAspGlySerSerAsnHisValTyrAsnTyr 520

```

Db      1558 GCACACTGCAGAAAGATACAGCTGAAAAAGACGGCTCCTTAACCATGTTTACACTAT 1617
Qy      521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
Db      1618 CAACCTGTCATCATCAGCGACGCTTGACAGTTCGGCTTGTCGATGACAA 1677
Qy      541 AamPheCysGluIlePheCysGlnCysSerSerGluCysGlnAaPheProGluCys 560
Db      1678 AATTTTGTGAAAAGTTTGTCTCAATGATGATTCAGATGTCMAACCGCTTCCGGGATGC 1737
Qy      561 ArgCysAlaAlaGlnCysAanthrIlyGlnCysProCysPheIleuAlaValArgGluCys 580
Db      1738 CGCTGCAGAACACAGTGCACACCAAGAGTGCCTGCTACTGCTGCTGCGAGAGTGT 1797
Qy      581 AapProAspLeuCysLeuThrCysGluAlaAlaAspHisTTrpAspSerIlyAsnValSer 600
Db      1798 GACCCCTGACCTGCTTACTGTCAGCGCTGCACCATTCGAGCAAGTAAATGTCCTC 1857
Qy      601 CysIlyAsnCysSerIleGlnArgIlySerIlyValHisIleuIleuAlaProSerArg 620
Db      1858 TGCAGAACTGCAGTATTCAGCGGGCTCCMAAAGCATCTATTGCTGGACCATCTGAC 1917
Qy      621 ValAlaGlyTTrpGlyIlePheIleIlyAspProValGlnIlyAsnGlnPheIleSerGlu 640
Db      1918 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCGTGCAGAAAATGAATTCATCTCAGAA 1977
Qy      641 TTrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyIlyValIlyTrpAspIly 660
Db      1978 TACTGGGAGAGATTTATTTCTCAGAGTGAAGCTGCAGAAAGGGAAAGTGATGATATA 2037
Qy      661 TyrMetCysSerPheLeuPheAsnIleuAsnAspPheValIleAspAlaThrArgIly 680
Db      2038 TACATGTGACGCTTCTGTTCACTTGAACAAATGATTTGTGTGATCAACCCGCAAG 2097
Qy      681 GlysAsnIlyIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaIlyValMet 700
Db      2098 GGTAACAAATTCGTTTTCATATCATTCGTAAATCCAAACTGCTATCAGAAAGTTAAG 2157
Qy      701 MetValAsnGlyAspHisArgIleGlyIlePheAlaIlyAsnGlnAlaIleGlnThrGlyGlu 720
Db      2158 ATGGTTAAGCGTATCAGAGTAGGTATTTTGGCAGAGAGCCATCCAGACTGGCGAA 2217
Qy      721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuIlyPheTyrValGlyIleGlu 740
Db      2218 GAGCTGTTTGTGATTTACATACAGCAGGCTGATGCCCTGAAGTATGTCCGCATCGAA 2277
Qy      741 ArgGluMetGluIlePro 746
Db      2278 AGAGAAATGGAATCCCT 2295

```

RESULT 8  
US-10-342-887-1043

```

; Sequence 1043, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14

```

```

; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1043
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-342-887-1043

Alignment Scores:
Pred. No.: 0
Score: 4038.00
Percent Similarity: 99.7%
Best Local Similarity: 99.7%
Query Match: 99.7%
DB: 8
Gaps: 0

US-10-773-302-2 (1-746) x US-10-342-887-1043 (1-2576)

Qy      1 MetGlyGlnThrGlyIlyValSerSerGluIlyValProValCysTrpArgIlyAspValIly 20
Db      58 ATGGGCCAGACTGGAGAAAGATCTGAGAGGAGCAAGTTTGTGGCGAAGCGTTAANA 117
Qy      21 SerGluTyrMetArgLeuArgGlnIlyValAspArgArgAlaAspGluValIlySer 40
Db      118 TCAGAGTACATCGACTGACAGACGCTCAAGAGTTCAAGCAGAGTGAAGTAAAGAGT 177
Qy      41 MetPheSerSerAsnArgGlnIlyIleGluGlnThrGluIleLeuAsnGlnIlyTrp 60
Db      178 ATGTTTAACTCCAAATCGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGATGG 237
Qy      61 LysGlnArgArgIleGlnProValHisIleuThrSerValSerSerLeuArgIlyThr 80
Db      238 AAACGCCAGAGATACAGCTGTGCAATCTGACTTCTGTAGCTCATTTGCCGGAGCT 297
Qy      81 ArgGluCysSerValThrSerAspLeuAspPheProthGlnValIleProLeuIlyThr 100
Db      298 AGGAGTGTTCGGTACCAAGTCACTGATTTTCCAAACAAGTATCATCCATTAAAGACT 357
Qy      101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnIlyAsnPhe 120
Db      358 CTGAATGAGATGCTTCAGTACCCATAATGATTTCTTGTCTCCCTCAGCAGAAATTT 417
Qy      121 MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAsp 140
Db      418 ATGTGTGAGAGTGAACCTGTTTACATACATCTCTTAAAGGAGATGAACTTTAGAT 477
Qy      141 GlnAspGlyThrPheIleGluIleuIleIlyAsnTyrAspGlyIlyValHisIleGlyAsp 160
Db      478 CAGATGTGATCTTTCATGAAGAACTAATAAATAATGATGGAAGTACACGGGAT 537
Qy      161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180
Db      538 AGAGAAATGTGGGTTTATTAATGAAGAAATTTTGTGGAGTTGTGAAGCCCTTGCTGCA 597
Qy      181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGlnGluGluIlyGln 200
Db      598 TATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
Qy      201 LysAspLeuGluAspHisArgAspAspIlyGluSerArgProProArgIlyPheProSer 220
Db      658 AAAAGTCTGAGAGTACCGAGATGATTAAGAAAGCCGCCACCTCGGAATTTCTTCT 717
Qy      221 AspIlyIlePheGluAlaIleSerSerMetPheProAspIlyGlyThrAlaGluGluLeu 240
Db      718 GATAAATTTTGGAGGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGACTA 777
Qy      241 LysGluIlyTyrIlyGluLeuThrGluGlnIlyLeuProGluValaLeuProProGluCys 260
Db      778 AAGGAAAATATTAAGAACTACCGAACAGAGCTCCAGGCGGACTTCTCTGAAATGT 837
Qy      261 ThrProAsnIleAspGlyProAsnAlaIlySerValGlnArgGluGlnSerLeuHisSer 280
Db      838 ACCCCCAACATGATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAGCAAGCTTACATCC 897

```

QY 281 PheH1eThrLeuPheCySarGArGcyPheLysTyraPcyPheLeuH1eProPheH1s 300  
Db 898 TTTCATACGCTTTCTGTGAGCGATGTTTAAATATGACTCTCTCATCATCTTTTCAT 957  
QY 301 AlaH1eProAenThrTyraGArGlybAsnThrguThraLaLeuAaPAsnLysProCyS 320  
Db 958 GCACACCCCAACACTTATAGCGAGAACACAGAACACTCTAGACACAAACCTTGT 1017  
QY 321 GlYProGlnCySyrGlnH1eLeuGlnGlyAlaYsGlnPheAlaAlaLeuThraLa 340  
Db 1018 GACACACAGCTTTACAGCATTTTGGAGGGACAAAGAGTTTGTGCTGCTCACCGCT 1077  
QY 341 GluArg11eLysThrProProLysArPProGlyGlyAlaGArGlyAlaLeuProAsn 360  
Db 1078 GAGCGGATAAAGACCCCAACAAACGTCAGAGGCGCCAGAGGGAGCGCTTCCCAT 1137  
QY 361 AsnSerSerArGProSerThrProThr11eAsnVal1eGlnUserLysArPThraPAser 380  
Db 1138 AACGATGACAGCCCAACGACCCCACTATATGCTGAGATCAAGATCAAGATCAAGACT 1197  
QY 381 AsPArGlnAlaGlyThrgLuthrgLuthrgLysLysLysLysLysLysLysLysLys 400  
Db 1198 GATAGGAGACAGGAGCTGAAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAG 1257  
QY 401 LysArPGLuthrSerSerSerSerSerGlnAlaAsnSerArGcyGlnThrPro11eLysMet 420  
Db 1258 AAAGATGAACCTTCAGAGCTCTCTGAAACAAATCTCCGTGCAACACATTAAGATG 1317  
QY 421 LysProAen11eGlnProProGlnAsnAlaGlnThraPAserGlyAlaGlnAlaSerMetPhe 440  
Db 1318 AAGCCAAATATTAAGACTCTCTGAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1377  
QY 441 ArgVal1eLeu11eGlyThrTyraTyraAsnPheCyAlaAlaLeuArgLeu11eGlyThr 460  
Db 1378 AGAGCTCATTTGGACACTTACATATCAATTTCTGTGCACTTGGATGATTAATGGAGCC 1437  
QY 461 LysThrCySarGlnVal1eGlnPheArgVal1eLysGlnUserSer11eLeuAlaProLa 480  
Db 1438 AAAACATGTAGCAGGTATATAGTTAGATCAAGATCAAGATCAAGATCAAGCTCACACT 1497  
QY 481 ProAlaGlnAsnVal1eAsPThrProProArgLysLysLysLysLysLysLysLysLys 500  
Db 1498 CCGCTGAGAGTGTGATATCTCTCAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGCT 1557  
QY 501 AlaH1eCySarGly11eGlnLeuLysLysArPGLysSerSerAsnH1eVal1eTyraAsnTy 520  
Db 1558 GCACACTGCAAGAAATATCAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1617  
QY 521 GlnProCySarPhe1eProArgGlnProCySarPAserSerCyProCyVal11eAlaGln 540  
Db 1618 CAACCTGTGATCATCCACAGCGAGAGCTTGTGACAGTTGCTGCTGCTGCTGCTGCTG 1677  
QY 541 AsnPheCyGlnLysPheCyGlnLysSerSerSerGlnCyGlnAsnArgPheProGlnCyS 560  
Db 1678 AATTTTGTGAAAGATTTTGTCAATGTATGTCAAGGTGCAAAACCGCTTCCGGAGAGC 1737  
QY 561 ArgCySlyAlaGlnCyAsnThrTyraGlnCySProCyTyraLeuAlaValArgLysCyS 580  
Db 1738 CGCTGCAAGACATGCAACCAAGAGAGTCCCGTGTACTGCTGCTGCTGCTGCTGCTGCT 1797  
QY 581 AsPProAsnLeuCyLeuThrCyGlnAlaAlaAsnH1eTTPAsPserLysAsnValSer 600  
Db 1798 GACCTGACCTCTGCTTATCTGTGAGAGCGCTGACCAATGGAGACATTAAGATATGCTCC 1857  
QY 601 CysLysAsnCySerSer11eGlnArgLysLysLysLysLysLysLysLysLysLysLys 620  
Db 1858 TGCAGAACTGCAATTCAGCGGAGCTCCAAAGAACATTAATGTGACACATCTGAC 1917  
QY 621 ValAlaGlyTTPGly11ePhe11eLysAsnProVal1eGlnLysAsnGlnPhe11eSerGln 640  
Db 1918 GTGGAGGCTGGGGATTTTATCAAGATCTCTGTGAGAGAGAGAGAGAGATTAATCAACA 1977  
QY 641 TyrCySgLyGln11eLysSerGlnAsnGlnAlaAsnArGArGlyLysVal1eTyraPAsy 660

Db 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAGAGAGAGATATGATTA 2037  
QY 661 TyraMetCySarPheLeuPheAsnLeuAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 680  
Db 2038 TACATGTGACAGCTTCTGTTCACTTGAACAATGATTTTGTGGATGAGCAACCGGAG 2097  
QY 681 GlnAsnLys11eArgPheAlaAsnH1eSerVal1eAsnProLysCyTyraAlaValMet 700  
Db 2098 GGTAAACAAATCTCTTGTGAAACATTCGATTAATCAACCTGTATGCAAAAGTTATG 2157  
QY 701 MetValAsnGlyAsnH1eArg11eGly11ePheAlaLysArgAla11eGlnThrgLysL 720  
Db 2158 ATGCTTAACGCTGATTCAGAGATGATATTTTCCAGAGAGCCATCCAGACTGGGAA 2217  
QY 721 GluLeuPhePheAsPtyraGtyrSerGlnAlaAsnAlaLeuLysTyraGly11eGln 740  
Db 2218 GAGCTGTTTGTGATTAACAATACAGCAAGCTGATCCCTGAAGTATGTGGCATCGAA 2277  
QY 741 ArgGlnMetGln11ePro 746  
Db 2278 AGAGAAATGGAATCCCT 2295

## RESULT 9

US-10-848-755A-129  
/ Sequence 129, Application US/10848755A  
/ Publication No. US2005054826A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Mao, Mao  
/ TITLE OF INVENTION: HUMAN DIAPHANOUS-3 GENE AND METHODS OF USE THEREFOR  
/ FILE REFERENCE: 9301-196-999  
/ CURRENT APPLICATION NUMBER: US/10/848, 755A  
/ PRIOR APPLICATION NUMBER: 60/471, 842  
/ PRIOR FILING DATE: 2003-05-11  
/ NUMBER OF SEQ ID NOS: 275  
/ SOFTWARE: PatentIn version 3.2 CAM: 301891-999188  
/ SEQ ID NO 129  
/ LENGTH: 2576  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-10-848-755A-129

Alignment Scores:  
Pred. No.: 0 Length: 2576  
Score: 4038.00 Matches: 744  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 2  
Query Match: 99.7% Indels: 0  
Gaps: 0  
DB: 10

US-10-773-302-2 (1-746) x US-10-848-755A-129 (1-2576)

QY 1 MetGlyGlnThrgLysLysSerGlnLysGlyProVal1eCyTTPAsyArGVal1e 20  
Db 58 ATGGGCGACACTGGAGAGAAATCTGAGAGAGACCAAGTTTGTGGCGAGACGCTATAA 117  
QY 21 SerGlnTyraMetArGLeuArgGlnLeuLysArgPheArgArgAlaAsnGlnVal1eSer 40  
Db 118 TCAAGATACATGCGACTGACAGCAAGTCAAGAGTTCAAGACAGCTGATGAAGTAAAGAGT 177  
QY 41 MetPheSerSerAsnAsnArgLysLys11eLeuGlnArgThrgLys11eLeuAsnGlnGly 60  
Db 178 ATGTTTACTTCATTCGTCAGAAATTTTGGAAAGAACGAAATCTTAACCAAGAAATG 237  
QY 61 LysGlnArgArg11eGlnProVal1e11eLeuThrSerVal1eSerSerLeuArgGlyThr 80  
Db 238 AAACAGCGAAGATACAGCTGTGACATCTGACTTTGTGAGCTCATTTGCGCGAGACT 297  
QY 81 ArgGlnCySerVal1eThrSerAsnLeuAsnPheProThrglnVal11eProLeuLysThr 100  
Db 298 AGGAGTGTGGGTGACCAAGTGAATTTTCCAAACAAAGTCATCCATTAAGACT 357

QY 101 LeuAnaLaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAspHe 120  
 DB |||||CTGAATGCGATTGCTTCACTACCCATTAATGATTTCTGGTCTCCCTTACAGCAAAATTTT 417  
 QY 121 MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAsp 140  
 DB 418 ATGGTGGAAAGATGAACCTGTTTATCACTAACTCCCTTAATATGGAGATGAGATGTTTAAAT 477  
 QY 141 GlnAspGlyThrPheIleGlnGluLeuIleIleValAsnTyrAspGlyLysValHisGlyIleAsp 160  
 DB 478 CAGGATGTGATCTTTCATTAAGAACTCAATTAATAATTAATGATGGAAAGATTAACCGGGAT 537  
 QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnGln 180  
 DB 538 AAGAGATGTGGGTTTATTAATATGAATTAATTTTGTGGAGTTGTGTGAATCCCTTGGTCA 597  
 QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluLysGlnGluLysGln 200  
 DB 598 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 QY 201 LysAspLeuGluAspHisAspAspAspLysGluSerAspProProArgLysPheProSer 220  
 DB 658 AAAGATCTGGAGATCAACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 717  
 QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGluLeu 240  
 DB 718 GATTAATAATTTGGAGGCAATTTCTCTCAATGTTCCAGATTAAGGGCAAGCAAGAAAGACTA 777  
 QY 241 LysGluLysTyrGlyGluLeuThrGlnGlnGlnLeuProGlyAlaIleuProProGluCys 260  
 DB 778 AAGAAAAAATTAAGAATCAACCAAGACAGCTCCAGGGCACTTCTCTGAAT 837  
 QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerIleHis 280  
 DB 838 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
 QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisAspPheHis 300  
 DB 898 TTTCATAGCTTTTCTGATGGCGATGTTTAATTAATGATGCTTCTCTCAATCTTTTCAAT 957  
 QY 301 AlaThrProAsnThrTyrTyrAspArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 DB 958 GCAACACCAACATTAATTAAGGGAAGAAACAAGAAACAGCTCTAGACAAACAACTTGT 1017  
 QY 321 GlyProGlnCysTyrGlnHisIleuGlnGlyAlaLysGluPheAlaAlaIleuThrAla 340  
 DB 1018 GGAACCAAGTGTATCAAGATTTGAGGAGGAGCAAGAGAGTTGCTGCTCTCAACCGCT 1077  
 QY 341 GlnArgIleLysThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360  
 DB 1078 GAGCGAATTAACACCCCAACAAACGTCACAGAGGCGCGAAGAGAGAGCGCTTCCCAAT 1137  
 QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleuGlnSerLysAspThrAspSer 380  
 DB 1138 AACATGACAGGCCCAAGCAACCCCAACATTAATGTGCTGATCAAGATCAAGACACT 1197  
 QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGlnAsnAspLysGluGlnGluLys 400  
 DB 1198 GATAGGAG 1257  
 QY 401 LysAspGluThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 420  
 DB 1258 AAAGATGAAACTTTCAGAGCTCTCTCAAGAGAAATTTCTGGTGTCAACCAATTAAGATG 1317  
 QY 421 LysProAsnIleGluProProGluLysValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 DB 1318 AAGCCAAATATTAACCTCTCAAGATGTGAGATGTGAGATGTGAGATGTGAGATGTGAGATGTG 1377  
 QY 441 ArgValIleuIleGlyThrTyrTyrAspAspPheCysAlaIleAlaArgIleuIleGlyThr 460  
 DB 1378 AAGATCTCAATGAGCACTTACTATGACAAATTTCTGTGCAATGCTAGAGTTAATGGGAC 1437  
 QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleIleAlaProAla 480

DB 1438 AAAACATGTAGACAGGTATATGAGTTTATGAGTCAAAAGATGTAGATCATAGCTCAAGCT 1497  
 QY 481 ProAlaGluAspValAspThrProProArgLysValLysValGlyValHisAspLeuThrAla 500  
 DB 1498 CCGCTGAGATGTGTGATATCTCTCAAGAAAGAAAGAGAGAAACACCGGTGTGGGGCT 1557  
 QY 501 AlaHisCysValArgValIleGlnLeuLysValAspGlyLysSerSerAsnIleValTyrAsnTyr 520  
 DB 1558 GCACCTCAAGAAATATCAAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 DB 1618 CAACCTGTATATCAACAGGAGCTGTGACAGTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1677  
 QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGluCys 560  
 DB 1678 AATTTTGTGAAAGTTTGTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
 QY 561 ArgCysValValGlnCysAsnThrLysGlnCysProCysTyrTyrLeuAlaValArgLysCys 580  
 DB 1738 CGCTGCAAGCAAGTGTCAACCAAGAGATGCCCCGTCTACCTGGCTGTCCGAGAGAT 1797  
 QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIlePheAspSerLysAsnValSer 600  
 DB 1798 GACCTGACCTCTGTCTTACTTGTGTGAGACCCCTGACATTTGGAGACATTAATAATGTGTCC 1857  
 QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleuLeuAlaProSerAsp 620  
 DB 1858 TGCAAGAACTCAAGATTCAGAGGGGCTCAAAAGCATGTATGCTGGCAACATCTGAC 1917  
 QY 621 ValAlaGlyTyrPheGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 DB 1918 GTGGAGGCTGGGGATTTTATCAAGATCTGTGCAAGAAAGATTAATCTCTCAGAA 1977  
 QY 641 TyrCysGlyGlnIleIleSerGlnAspGluAlaAspAspArgGlyLysValTyrAspLys 660  
 DB 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAAAGAGAGAAAGTGTATGATAA 2037  
 QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
 DB 2038 TACATGTCACTTTCTGTCTTCAACTGAACATGATTTTGTGTGATGCAACCCGCAAG 2097  
 QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
 DB 2098 GGTAACAAAATTCGTTTGAATCAATCTGTTAATCAAACTGCTATGCAAAAGTTATG 2157  
 QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
 DB 2158 ATGTTTAACGGTATCAAGATAGATATTTTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2217  
 QY 721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGlu 740  
 DB 2218 GAGCTGTTTGTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 2277  
 QY 741 ArgIleMetGluIlePro 746  
 DB 2278 AGAGAAATGAAATCCCT 2295  
 RESULT 10  
 US-10-504-173-78  
 ; Sequence 78, Application US/10504173  
 ; Publication No. US2005020428A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Axiordia Limited  
 ; TITLE OF INVENTION: Pluripotential Stem Cells  
 ; FILE REFERENCE: P101863WO  
 ; CURRENT APPLICATION NUMBER: US/10/504,173  
 ; PRIOR APPLICATION NUMBER: 0203359.5  
 ; PRIOR FILING DATE: 2002-02-02  
 ; NUMBER OF SEQ ID NOS: 135  
 ; SOFTWARE: PatentIn version 3.1

SEQ ID NO 78  
LENGTH: 2576  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-504-173-78

## Alignment Scores:

Pred. No.:	0	Length:	2576
Score:	4038.00	Matches:	744
Percent Similarity:	99.7%	Conservative:	0
Best Local Similarity:	99.7%	Mismatches:	2
Query Match:	99.7%	Indels:	0
	10	Gaps:	0

US-10-773-302-2 (1-746) x US-10-504-173-78 (1-2576)

QY 1 MetGlyGlnThrGlyLysLeuSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
DB 58 ATGGGCCAGACTGGAGAAATCTGAGAAAGGAGCCAGTTGTGGCGGAGCGGTAAAA 117  
QY 21 SerGlyTrpMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 118 TCAGAGTACATGCCGACTGAGCAGCTCAAGAGTTCAAGCAGCTGATGAAGTAAAGCT 177  
QY 41 MetPheSerSerAspAspArgGlnLysIleLeuGluValArgThrGluIleLeuAsnGlnGluTrp 60  
DB 178 ATGTTAGTCCAAATCTCAGAAAATTTGGAAAGACGAAATCTTAAACCAAGAAATGG 237  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 238 AAACGCGAAGGATACAGCTGTGCAATCTGACTTCTGTAGCTCATTTGGCGGGACT 297  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 298 AGGAGGTCTTCGTGACCAAGTCACTTGGATTTCCAAACAAGTATCCCATTAAGAATCT 357  
QY 101 LeuAsnAlaValAlaSerValProIleMetCysTrpSerProLeuGlnAsnPhe 120  
DB 358 CTGAATGCAGTTCCTCAGTACCCCAATATGTAATTCCTGCTCCCTCAGACGAAATTTT 417  
QY 121 MetValGluAspGlnThrValIleuHisAsnIleProCysTrpMetGluAspGluValLeuAsp 140  
DB 418 ATGGGGAAGAGTAACTGTTTACATACATTCCTTATATGGGAGATGAAGTTTAAAT 477  
QY 141 GlnAspGlyThrPheIleGlnGluLeuLeuLysAsnIleAspGlyLysValHisGlyAsp 160  
DB 478 CAGAGTGTACTTTCATTGGAAGACTTAATTAAGATTAATGATGGAAAGTACACGGGAT 537  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 538 AGAGAAATGTGGGTTTAAATGATGAAATTTTGTGGAGTTGTGAATGCCCTTGTGCTCA 597  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnGluLysGln 200  
DB 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
QY 201 LysAspLeuGluAspHisAlaArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 658 AAAGATCTGGAGAGTACCGGAGATGATTAAGAAAGCCGCCACCTCGGAATTTCTCTCT 717  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGluLeu 240  
DB 718 GATTAATTTTGGAGCGCATTTCTCAATGTTTCCAGATTAAGGGGACACGCAAAAGACTA 777  
QY 241 LysGluLysTrpLysGluLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
DB 778 AAGGAAAAATATTAAGAACTCACCGAAGAGCTCCAGCGGCACTTCTCTCTTAAGT 837  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer 280  
DB 838 ACCCCCAACATGATGAGCAACAAATGCTAAATCTGTTCAAGAGACAAAGCTTACATCC 897  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300

DB 898 TTTCATACGCTTTCTGTAGCCATGTTTAAATATATACCTCTTCAATCTTTTCAT 957  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB 958 GCACACCCCAACATCTTAATTAAGCCGAGAAACAGAAACAGCTTACCAACAACTTGT 1017  
QY 321 GlyProGlnCysTyrGlnHisLeuGlnGluValAlaLysGluPheAlaAlaIleThrAla 340  
DB 1018 GCACCAAGTGTTCACGACTTTGGAGGAGCAAGAGATTTGTCTGCTCTACCGCT 1077  
QY 341 GluArgGlnLysThrProProLysArgProGlyValArgArgArgGlyArgSerProAsn 360  
DB 1078 GAGCGGATTAAGACCCCAACAAAGCTCAGAGAGCGCCGACAGAGAGAGCGCTTCCAAAT 1137  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1138 AAGAGTGAAGGCCCAAGACACCCCACTTAATGTGCTGAATCAAGATACAGACAGT 1197  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyValAsnAspAspLysGlnGlnGlnLys 400  
DB 1198 GATAGGAGAGCAGGAGCTGAACCGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1257  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1258 AAAGATGAACCTTGAGCTCTCTGAGCANAATCTCGGTCTCAACACCAATTAAGATG 1317  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1318 AAGCAATATATTGAACCTCTGAGAAATGTGAGAGTGAGTGCTGAAGCTCCATATGTT 1377  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaGlyLeuIleGlyThr 460  
DB 1378 AGAGTCTCATTTGACACTTACTATGACAATTTGTGTCCATGTCTAGGTTAATGGAGCC 1437  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB 1438 AAACATGTATACAGGTGTATGATGATTTAGATCAAAAGATTTACATACATGCTCCAGCT 1497  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
DB 1498 CCGCTGAGATGTGATGATCTCTCAAGAAAGAAAGAAAGAAAGCAACCGGTGTGGGCT 1557  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnThr 520  
DB 1558 GCACACTGCAAGAAAGTACAGCTGAAGAAAGAGCGGCTCTCTTAACATGTTTACACTAT 1617  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1618 CAACCTGTATCTTCCAGCGGACGCTTGTGACAGTTCGCTCTGTGTGATAGCAAA 1677  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
DB 1678 AATTTTGTGAAGATTTTGTCAATGTATGTCAAGATGTCAAAACCGCTTCCGGGATGC 1737  
QY 561 ArgCysLysValGlnCysAsnThrLysGlnCysProCysTyrTrpAlaAlaValArgGlnCys 580  
DB 1738 CGCTGCAAGACAGGTGCAACCAAGCAGTGGCCGTGTCACTCGGCTGTCCAGAGAGT 1797  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1798 GACCTGACCTCTGTCTTACTTGTGAGACCGCTGACCACTTGGACAGTAAATATGTGCTC 1857  
QY 601 CysLysAsnCysSerIleGlnArgLysSerLysHisLeuLeuLeuAlaProSerAsp 620  
DB 1858 TGCAAGAACTCAGATATCAACGGGGCTCCAAAAGCATATATGCTGGACCATCTGAC 1917  
QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1918 GTGGAGGCTGGGGATTTTATCAAAATCTGTGCGAAGAAATGAATTCATCTCAGAA 1977  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660





Db	1438	AAAAAAGTAAACAGGTGTATGAGTTTAAAGTCAAAAGATTAGATCATAGCTCAAGCT	1497
Qy	481	Pro1aGluAerValAspThrProAArgLysLysValArgLysHisArgLeuTrpAla	500
Db	1498	CCCGCTGAGATGTGGATATCTCTCCAAAGAAAAAGAAAGAAACACCGGTGTGGGCT	1557
Qy	501	AlaHisCysValArgLysLLeGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr	520
Db	1558	GCACACGTGCAGAAAAGATACAGCTGAAAAAGACCGGCTCTCTTAACCAAGTTTACACATAT	1617
Qy	521	GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln	540
Db	1618	CAACCCGTGTATCATTCACGGCAGCCTTGTACAGATTGTGACCTTGTTGTATAGCACAA	1677
Qy	541	AsnAspCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys	560
Db	1678	AAATTTTGTAAAAAGTTTGTGCATGTATGTTCAAGGTGTCAAAAACCGTTTTCCGGAGTC	1737
Qy	561	ArgCysLysValaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys	580
Db	1738	CGCTGCAGAAAGCACAGTGCAGAACCAACAGAGTCCCGGTCTACCTGGCTGTCCGAGAGTGT	1797
Qy	581	AspProAspLeuCysLeuThrCysGlnValAlaAspHisTrpAspSerLysAsnValSer	600
Db	1798	GACCCCTGACCTCTGTCTTACTTGTGTGACCCGCTGCACATGTGGAGAGTAAAAATGTGTCC	1857
Qy	601	CysLysAsnCysSerLLeGlnArgGlySerLysLysHisLeuLeuLeuAlaProSerAsp	620
Db	1858	TGCAGAGATCTCAGATATTCAAGGGGGCTCCAAAAAGCATCTATTGTGGACATCTGAC	1917
Qy	621	ValAlaGlyTrpGlyTyrPheTyrLeuAspProValGlnLysAsnGluPheLleSerGlu	640
Db	1918	GTGGCAGGCTGGGGGATTTTATCAAGATGACGTGCAGAAAATGAAATCATCTCAAGAA	1977
Qy	641	TyrCysGlnGlyLulleLleSerGlnAspGlnAlaAspArgGlyLysValTyrAspLys	660
Db	1978	TACTGTGAGAGATTATTCTCAAGATGACGTGCAGAAAAGGAAAGTGTATGATAAA	2037
Qy	661	TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys	680
Db	2038	TRCATGTGCACTTCTCTTCACTGTGAACATGATTTTGTGGTGGAGTGAACCCGCAAG	2097
Qy	681	GlyAsnLysLLeArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet	700
Db	2098	GGTAAACAATAATTCGTTTGGCAAAATCATTCGGTAAATCCAAACCTGCTATGCAAAAGTTATG	2157
Qy	701	MetValaAsnGlyAspHisArgGlyLLeGlyLLePheAlaLysValaGlnIleGlnThrGlyGlu	720
Db	2158	ATGTGTAAACGGTGAACACAGATAGATGATTTTGTGCAGAGAGCCATCAACATGGCCGAA	2217
Qy	721	GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuLysTyrValGlyLLeGlu	740
Db	2218	GAGCTGTTGTGTGATTACAGATACACCAAGGCTGATGCCCTGAGATATGTGGCATGAA	2277
Qy	741	ArgGluMetGluLeuPro 746	
Db	2278	AGAGAAATGAAATCCCT 2295	
RESULT 12			
US-10-210-120-95			
; Sequence 95, Application US/10210120			
; Publication No. US20030175736A1			
; GENERAL INFORMATION:			
; APPLICANT: Chinmayan, Arul M.			
; APPLICANT: Rubin, Mark A.			
; APPLICANT: Steekumar, Arun			
; TITLE OF INVENTION: Expression Profile of Prostate Cancer			
; FILE REFERENCE: UM-07221			
; CURRENT APPLICATION NUMBER: US/10/210,120			
; CURRENT FILING DATE: 2002-08-01			
; PRIOR APPLICATION NUMBER: US 60/309,581			
; PRIOR FILING DATE: 2001-08-02			

PRIORITY APPLICATION NUMBER: US 60/334,468									
PRIOR FILING DATE: 2001-11-15									
NUMBER OF SEQ ID NOS: 123									
SOFTWARE: PatentIn version 3.2									
SEQ ID NO 95									
LENGTH: 2512									
TYPE: DNA									
ORGANISM: Homo sapiens									
US-10-210-120-95									
Alignment Scores:									
Pred. No.:	0	Length:	2512						
Score:	4024.00	Matches:	745						
Percent Similarity:	99.94	Conservative:	0						
Best Local Similarity:	99.98	Mismatches:	1						
Query Match:	99.34	Indels:	1						
DB:	7	Gaps:	0						
US-10-773-302-2 (1-746) x US-10-210-120-95 (1-2512)									
QY	1	MecGlyGlnThrGlyLysLysSerGluLysGlyProValCysThrArgLysArgValLys	20						
DB	10	ATGGGCCAGACTGGGAGAGAAATCTGAGAAAGGACCAAGTTGTTCGGAGGAGCTGTAA	69						
QY	21	SerGluYrMetCArgLysArgGlnLysLysArgPheArgArgAlaAspGluValLysSer	40						
DB	70	TCAAGTACATGCGACTGACAGAGCTCCAGAGGTTCAAGCGAGCTGATGAAGTAAAGC-	128						
QY	41	MetPheSerSerAsnArgGlnLysLysLeuGluArgThrGluLysLeuAsnGlnGluTrp	60						
DB	129	ATGTTTATTCCTCAATCGTCAGAAATTTTGGAGAAAGACGGAATCTTAAACAGATGG	188						
QY	61	LysGlnArgArgLysGlnProValHisLysLeuThrSerValSerSerLeuArgLysThr	80						
DB	189	AAACGCGAGAAAGATCACCTGTGCACATCTCGACTTCTGAGGCTCATTTGCGGGAGCT	248						
QY	81	ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValLysProLeuLysThr	100						
DB	249	AGGAGGTGTGGTGACCAAGTACTGGATTTTCCACACAGTCAATCCCATTAAGACT	308						
QY	101	LeuAsnAlaValAlaSerValProLysMetYrSerTrpSerProLeuGlnGlnAsnPro	120						
DB	309	CTGATGACAGTTGCTTCACTGATGCCAATTAATGATTTCTTGGTCTCCCTCAAGCAGATTTT	368						
QY	121	MetValGluAspGluThrValLeuHisAsnLysLeuProYrMetGlyAspGluValLeuAsp	140						
DB	369	ATGGTGGAGAAAGTAAACGTTTACATTAACATTCCTTAATGAGATGAAGTTTAAAGT	428						
QY	141	GlnAspGlyThrPheLysGlnLysLeuLysAsnYrAspGlyLysValHisGlyAsp	160						
DB	429	CAGAGTGTACTTCAATTGAGAAACATAAATAAATAATTAAGTGGAAAGTACCGGGAGT	488						
QY	161	ArgGluCysGlyPheLysAsnAspGluLysPheValGlnLeuValAsnAlaLeuGluGln	180						
DB	489	AGAGATGTGGTTTATTAATGATGAATTTTGTGGAGTTGTGATGCCCTTGGTCAA	548						
QY	181	TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln	200						
DB	549	TATATATGATGATGACGATGATGATGAGACGATCTCGAAGAAAGAGAAAGAAAGCAG	608						
QY	201	LysAspLeuGluAsnHisArgAspAspLysGlnSerArgProProArgLysPheProSer	220						
DB	609	AAAGATTCGAGGATCCACGAGATATTAAGAAAGCGCGCACCTCGGAAATTTCTTCT	668						
QY	221	AspLysLysLysPheGlnAlaLysSerSerMetPheProAspLysGlyThrLysGlnGln	240						
DB	669	GATTAATATTTTGAAGCCATTTCTCTCAATGTTTCCAGTAAAGGGCACACAGAGAACTA	728						
QY	241	LysGluLysYrLysGlnLeuThrGlnGlnGlnLeuProGlyValLeuProProGluCys	260						
DB	729	AAGGAAATATTAAGAACTACCGAAACGACGCTCCGAGCGCACTTCTCTCGATATG	788						
QY	261	ThrProAsnLysAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer	280						

```

DB 769 ACCCCACATGATGAGCAAAATCTGTTACAGAGAGCAAAAGCTTACATCTC 848
QY 281 Pheh1srhLeupheCyarghargCyphelysTyaspCyaphelunhiaprophelis 300
DB 849 TTTCTACAGCTTTTCTGTAGCGAGATGTTTAAATATGATGCTTCTCATCATCTTTTCAT 908
QY 301 AATATrProAntThrTyrIlysharglyshAntThrgIuThAlaleuAspAntlyserProCys 320
DB 909 GCAACACCCACATTAATAGCGAAGAACACAGAAACGCTCTGACACAAACCTTGT 968
QY 321 GIPProGInCyTyrgInhiIstleugIuAlalySGluPheAlaAlaleuthrala 340
DB 969 GGAACACAGCTTACAGATTTGAGAGGAGCAAAAGATTGCTGCTCTCTCACCGCT 1028
QY 341 GIuArgIlelyeThrProProIyshaargProGIyIyargargargGIyArgleuProshn 360
DB 1029 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGGCGCGAGAAAGAGACGGCTCCCAT 1088
QY 361 AsenSerSerArgProSerThrProThrlleAntValleuGIuSerIlyAspThrAspSer 380
DB 1089 AACAGTAGAGGCGCCAGACACCCCAACATTAATGTGTGAATCAAAAGGATACAGACGT 1148
QY 381 AsparGluAlaGIyThrgIuThrgIyGIuAsnAsnAspIySGluGIuGIuIyus 400
DB 1149 GATAGGAGAGAGGAGATGAAACGGGGGAGAGAAACATGATTAAGAGAGAGAGAGANG 1208
QY 401 LyAspGIuThrSerSerSerSerGIuAlaAsnSerArgCySGInThrProIleIysMet 420
DB 1209 AAAGATGAACCTTCGAGGCTCCTCTGAGGAATTCCTCGGTGCAAAACCATTAAGATG 1268
QY 421 LysProAntIleGIuProProGIuAntValGIuTPSerGIyAlaGIuAlaSerMetPhe 440
DB 1269 AAGCAAAATATTTGAACCTCTGAGATGTGGAGGTGGAGGTGCGAAGCCCAATGTTT 1328
QY 441 ArgValleuIleGIyThrTyrTyraAsnPheCyAlaIleAlaArgleuIleGIyThr 460
DB 1329 AGAGTCTCATTTGGGACTTACTATGACAAATTTCTGTGCTTGTAGGTTAATTTGGACC 1388
QY 461 LysThrCyargGIuValIyrgIuPheArgValIySGIuSerSerIleIleAlaProAla 480
DB 1389 AAAACATGTAGACAGGTGTATGATTAAGTCAAAAGATCTAGCATCATCTCAGCT 1448
QY 481 ProAlaGIuAspValAspThrProProArgIyGIyIyAlaGIyGhiIeArgleuTPAla 500
DB 1449 CCGGCTGAGAGTGTGATCTCTCCAAAGAAAGAAAGAGAAACACCGGTGTGGGCT 1508
QY 501 AlaHISCyAsrgIyAlleGIuIleuIyIyAspGIySerSerAsnHISValIyTrAsnTyr 520
DB 1509 GACACCTGAGAAAGATACAGCTGAAGAAAGAGCGGCTCTCTTAACCATGTTTACAACTAT 1568
QY 521 GInProCyAspPheAspArgIuProCyAspSerSerCyPProCyValIleAlaGIn 540
DB 1569 CAACCTGTGATCATCCAGGAGGCTTGTGACATGTGGTGCCTTGTGTGATACACAA 1628
QY 541 AsnPheCySGIuIyPheCySGInCySserSerGIuCySGInAsnArgPheProGIyCys 560
DB 1629 AATTTTGTGAAGATTTTGTCAATGTATGTCAAGGTGCACAAACCGCTTCCGGGAGTCC 1688
QY 561 ArgCyIyAlaGIuCyAsnThrIySGInCySPProCyIyIyLeuAlaValaArgIuIyus 580
DB 1689 CGCTGCAAGGACAGGTGACACCAAGAGGTGCCGTGTACTGTGCTGTCCGAAGGTGT 1748
QY 581 AspProAspLeuCyIeuthrCySGIyAlaAlaAspPheIstTPAspSerIyshAntValSer 600
DB 1749 GACCTGAGCTCTGTCTTACTTGTGAGCGCTGACCATTTGGAGACATTAAGATGTGTCC 1808
QY 601 CysIyAsnCySerIleGIuArgIySerIyIyGhiIeIleuIleuAlaProSerAsp 620
DB 1809 TGCAGAAACTGACATTTCAAGCGGCGCTCCAAAGAGCATCTATTCTGCGCACATCTGAC 1868
QY 621 ValAlaGIyTTPGIyIlePheIleIyAspProValGIuIyAsnGIuPheIleSerGIu 640

```

```

DB 1869 GTGGAGGCTGGGGATTTTATTCAAAGATCCTGTGACAGAAATGAATTCATCTCAGAA 1928
QY 641 TyrCySGIyGIuIleIleSerGIuAspGIuAlaAspArgargGIyIyValIyTrAspIyS 660
DB 1929 TACTGTGAGAGATTAATTTTCAAGATGAAGCTGACAAAGGAGAAAGTGTATATATAA 1988
QY 661 TyrMetCySerPheIleuPheAsnIleuAsnAspPheValIleAspAlaThrArgIyS 680
DB 1989 TACATGTCAAGCTTTCTGTTCATCTTGACATTAATGATTTTGTGTGATGCAACCGGANG 2048
QY 681 GIyAsnIyIleargPheAlaAsnHISerValAsnProAsnCyTyrgAlalyValMet 700
DB 2049 GGTAACAAAATTCGTTTGGCAATCATTCGGTAAATCCAACTGCTATGCAAAAGTTATG 2108
QY 701 MetValaenGIyAspPheArgIleGIyIlePheAlalyshargAlaIleGIuThrgIyGIu 720
DB 2109 ATGTTTAAACGGTGAACACAGATAGGTATTTTGTCCAAAGAGGCTATCAGCTGGCAGA 2168
QY 721 GIuIeuphePheAspTyrArgTyrSerGIuAlaAspAlaIeIyIyIleGIu 740
DB 2169 GAGCTGTTTGTGATTAACATACAGATACAGGCTGATGCTGAGATATGTGGCATCGAA 2228
QY 741 ArgGIuMetGIuIlePro 746
DB 2229 AGAGAAATGAATAATCCCT 2246

```

## RESULT 13

US-10-909-035-95  
Sequence 95, Application US/10909035  
Publication No. US20050136493A1

GENERAL INFORMATION:  
APPLICANT: Rubin, Mark A.  
APPLICANT: Chinaiyan, Arul M.  
APPLICANT: Lakman, Bharathi  
APPLICANT: Sreekumar, Arun

TITLE OF INVENTION: AMACR Cancer Markers  
FILE REFERENCE: IM-09098  
CURRENT APPLICATION NUMBER: US/10/909,035

NUMBER OF SEQ ID NOS: 123  
NUMBER OF SEQ ID NOS: 2004-07-30  
SOFTWARE: PatentIn version 3.2

SEQ ID NO 95  
LENGTH: 2512  
TYPE: DNA  
ORGANISM: Homo sapiens

US-10-909-035-95

Alignment Scores:  
Pred. No.: 0 Length: 2512  
Score: 4024.00 Matches: 745  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.3% Indels: 1  
DB: 10 Gap#: 0

US-10-773-302-2 (1-746) x US-10-909-035-95 (1-2512)

```

QY 1 MetGIyInThrgIyIyIySserGIuIySGIyProValCySTrPArgIyshArgValIyS 20
DB 10 ATGGGCCAGACTGGGAAATAATCTGAGAAAGGAGCAAGTTTGTGGCGAAGGTGTAATAA 69
QY 21 SerGIuIyMetArgIeArgIeGIuIleuIyArgPheArgAlaAspGIuValIySser 40
DB 70 TCAGAGTACAGGCACTGACAGACAGTCAAGAGGTTCACACAGAGCTGATGAAGTAAAGGT- 128
QY 41 MetPheSerSerAsnArgGIuIyIleuGIuArgThrgIuIleuAsnGIuGIuIy 60
DB 129 ATGTTAGTTCATTCGTCAGAAATTTTGAAGAGAGGAAATCTTAACCAAGAAATGG 188
QY 61 LysGIuArgArgIleGIuProValHISIleuThrSerValIserSerIeArgGIyThr 80
DB 189 AAACAGGAGAGATACAGCTGTGACATCTGACTTGTGTGAGCTATTGCGGCGGACT 248

```

QY ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100  
Db AGGAGAGTTCGGTGAACAGTGAAGCTTTCACACACAAAGTCAATCCCATTAAGACT 308  
QY LeuAsnAlaValAlaSerValProIleMetTyrSerTyrProLeuGlnIleAspPhe 120  
Db CTGAATGCAAGTTCCTGAGTACCAATATGTAATCTTGCTCCCTACAGCAAAATTTT 368  
QY MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAsp 140  
Db ATGTGTGAAGATGAAGACGTTTTCATTAACATTCCTTATATGGAGATGAAGTTTAGAT 428  
QY GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
Db CAGATGGTACTTCATGGAAGAACTAATAAAATTAATGATGGGAAAGTACAGGGGAT 488  
QY ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaIleuGlyGln 180  
Db AGAAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTGTGGAATGCCCTTGCTCAA 548  
QY TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
Db TATATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608  
QY LysAspLeuGluAspPheSarAspAspLysGluSerArgProArgGlyLysPheProSer 220  
Db AAAGATCTGGAGATCACCGAGATGATGAAGAAAGCCGCCCATCCGGAATTTCTTCT 668  
QY AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
Db GATAAATTTTGTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGACACGCAAGAAACTA 728  
QY LysGluLysTyrLysGluLeuThrGlnGlnIleuProGlyAlaIleuProProGluLys 260  
Db AAGGAAAAATTAAGAAGCTACCGAAGACAGCTCCAGCGCGCATTCCTCGAATGT 788  
QY ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnIleuHisSer 280  
Db ACCCCAAATATAGTGAACAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACATCC 848  
QY PheIleThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
Db TTTCATACGCTTTTCTGTAGCGATGTTTAATATGATGCTCTTCAATCCCTTTTAT 908  
QY AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaIleuAspAsnLysProCys 320  
Db GCACACCCCAACACTTAATACCGAAGAACACAGAAACAGCTTAGACAAACCTTGT 968  
QY GlyProGlnCysTyrGlnHisIleuGluGlyAlaLysGluPheAlaAlaIleuThrAla 340  
Db GGACCAACAGTGTACAGCATTTTGGAGGAGCAAAAGAGTTTGGCTGCTCTCACCCCT 1028  
QY GlnArgGlyLeuThrProProLysArgProGlyLysArgArgArgGlyArgLeuProAsn 360  
Db GAGCGAATAAAGACCCCAACAAACGTCAGAGAGCCGCAAGAGAGGAGGCTTCCAT 1088  
QY AsnSerSerArgProSerThrProThrIleAsnValIleuGluSerTyrAspThrAspSer 380  
Db AACGTAGCAAGGCCCAAGCAACCCCAACATTAATGCTGGAATCAAGATACAGACGT 1148  
QY AspArgGluAlaGlyThrGluThrGlyGlyLysAsnAspLysGluGluGluLys 400  
Db GATAGGGAAGCAAGGATGAAGAACGGGGGAGAGAAACAATGATTAAGAAAGAAAGAG 1208  
QY LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleMet 420  
Db AAAATGTAACCTTCAGGCTCTCTGAAACAAATTTCTGCTGTCAAAACCAATTAAGATG 1268  
QY LysProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
Db AAGCCAAATATTAAGCTCTCTGAAGATGAGTGAATGTGTGTGAAGCCCTCAATGTTT 1328  
QY ArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460

Db AGAGTCTCATGGACATTAATCTATGACAAATTTCTGTGCCATGTGATGATTAATGGAC 1388  
QY LysThrCysArgGlnValTyrGluPheArgValLysGluSerIleIleAlaProAla 480  
Db AAAACATGTACAGAGTGTATGAGTTAGAGTCAAAAGATCTAGCATATAGCTCCAGCT 1448  
QY ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
Db CCCGTGAGGATGTGATACTCTCTCCAAAGAAAGAAAGAGAAACCCGGTTGTGGGCT 1508  
QY AlaHisCysArgGlyIleGlnLeuLysLysAspGlySerSerSerHisValTyrAsnTyr 520  
Db GCACATGTCAGAAAGATACAGCTGAAGAAAGAGAGGAGCTCTTAACATGTTTACACTAT 1588  
QY GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db CAACCTGTATCATCAACGAGCGCTTGTGAAGTTGATGCGCTTGTGTATGACAA 1628  
QY AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
Db AATTTTGTGAAAAATTTGTCAATGATGATGATGATGATGATGATGATGATGATGATG 1688  
QY ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
Db CGTGCAGAGACAGTCAACACCAACAGAGTCCCGTCTACTCTGCTCCAGAGTGT 1748  
QY AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisThrAspSerLysAsnValSer 600  
Db GACCTGACCTCTGTCTTACTGTGTGAGCGCGTGAACATTTGGGACATTAATAATGTCTC 1808  
QY CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleuLeuLeuAlaProSerAsp 620  
Db TGCAAGAACTGAGTATTCAGCGGGGCTCCAAAGAAAGCATATTTGCTGGCAACATCTGAC 1868  
QY ValAlaGlyThrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db GTGGCAGCGTGGGGGATTTTATCAAGATCTGTGCAAGAAATGAAATTCATCTCGAA 1928  
QY TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
Db TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTATGATATA 1988  
QY TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
Db TACATGTGCGCTTTCTGTTCAACTTGAACATATATTTGTGGTGAATGCAACCCGCAAG 2048  
QY GlnAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysValMet 700  
Db GGTAACAAAATTCGTTTGGCAATCATTCGGTAAATCCAAATCTGCTAAGAAAGTTATG 2108  
QY MetValAsnGlyAspHisArgGlyIleGlyIlePheAlaLysArgAlaIleGlnThrGlyGlu 720  
Db ATGTGTAAAGGTATGATCAAGATAGATATTTTGTCCAAAGAGACCATTCAGACTGGGAA 2168  
QY GlnLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaIleuLysTyrValGlyIleGlu 740  
Db GAGCTGTTTTTGTATCAAGTACAGCAGGCTGATGCCCTGAAGTATGTGCGATCGAA 2228  
QY ArgGluMetGluIlePro 746  
Db AGAAGAAATGGAAATCCCT 2246

RESULT 14  
US-10-104-047-1192  
; Sequence 1192, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: HL-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25

PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1192  
LENGTH: 2476  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-047-1192

Alignment Scores:  
Pred. No.: 0 Length: 2476  
Score: 3812.50 Matches: 706  
Percent Similarity: 94.64 Conservative: 0  
Best Local Similarity: 94.64 Mismatches: 1  
Query Match: 94.14 Indels: 39  
DB: Gaps: 1  
US-10-773-302-2 (1-746) x US-10-104-047-1192 (1-2476)

QY 1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
DB ATGGCCAGACTGGGAGAGAAATCTGAGAGGAGCCAGTTGTGGCGGAGCGTAAATA 152  
QY 21 SerGlyTrpMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB TCAGAGTCAATGCGACTGAGACAGCTCAAGAGTTCAAGCAGCTGATGAGATAATAGT 212  
QY 41 MetPheSerSerAspArgGlnLysIleLeuGlnArgTrpGlnIleLeuAsnGlnLysTrp 60  
DB ATGTTTAATTCACATCGTCAGAAATTTTGGAAAGAGGAAATCTTAAACCAAGAAATG 272  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB AAACGCCAGAGATACAGCTGTGCAATCTGACTTCGTGAGCTCATTCGCGGAGACT 332  
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB AGGGAG----- 338  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnAsnPhe 120  
DB ----- 338  
QY 121 MetValGluAspGlnThrValIleuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB ----- 339  
QY 141 GlnAspGlyThrPheIleGlnLysLeuIleLysAsnTyrAspGlyLysValHisIleLysAsp 160  
DB ----- 339  
QY 161 ArgGlyCysGlyPheIleAsnAspGlnIlePheValGluLeuValAsnAlaLeuGlnLysGln 180  
DB AGAGATGTGGTGTATTAATGATGAATTTTGTGAGTTGTGTAATGCCCTTGTCTCA 515  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnLysArgGlnLysGln 200  
DB TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575  
QY 201 LysAspLeuGlnLysAsnHisArgAspAspLysGlnLysAsnArgProProArgLysPheProSer 220  
DB AAAGATCTGAGAGATCACCGAGATGATAAGAAACCCGCCCTCGGAAATTTCTCTCT 635  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLysLys 240  
DB GATATAATTTTGAAGCCATTTCTCATGTTTCCAGTTAAGGAGACACGAGAAAGACTA 695  
QY 241 LysGlnLysTyrLysGlnLysLysThrGlnLysGlnLysProGlyValAlaLeuProProGlyCys 260  
DB AAGGAAATATATAAGAACTCACCGAAGCAGCTCCAGGCGCACTTCTCTGAATGT 755  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnLysSerLeuHisSer 280

DB 756 ACCCCAAATATATGAGCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACATCC 815  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB TTTCATAGCGCTTTCTGAGGCGATGTTTTAAATATGATCTCTTCAATCTTTTCAT 875  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB GCACACCCAACTTATATAGCGGAGAGAACACAGAACAGCTTATAGACAAACCTTGT 935  
QY 321 GlnProGlnCysTyrGlnHisLeuGlnLysAlaLysGlnPheAlaAlaAlaLeuThrAla 340  
DB GGACCAAGCTTTTACAGCACTTTGGAGGAGCAAAAGAGTTTGTGCTGCTCAACGCT 995  
QY 341 GlnArgGlnLysThrProProLysArgProGlyGlyValArgArgGlyValArgLeuProAsn 360  
DB GAGCGATTAAGAACCCCAACCAACGTCACAGAGCGCCAGAGAGGAGCGGCTTCCCAT 1055  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnLysAspThrAspSer 380  
DB AACAGTACAGGCCCAAGCAACCCCACTTATATGTGCTGGAATCAAGATACAGACAGT 1115  
QY 381 AspArgGlnAlaGlyThrGlnLysGlyLysAsnAspLysGlnLysGlnLysLys 400  
DB GATAGGAG 1175  
QY 401 LysAspGlnThrSerSerSerSerSerGlnLysAsnSerArgCysGlnThrProIleLysMet 420  
DB AAAGATGAATCTTGAGGCTCTGAGAGCAATCTCGGTCTCAACACCAATTAAGATG 1235  
QY 421 LysProAsnIleGlnProProGlnAsnValGlnTrpSerArgValAlaGlnAlaSerMetPhe 440  
DB AAGCAATATATTAACCTCTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB AGAGTCTTCATTTGACCTTATATGACAAATTTCTGTGCACTTGTGATTTATTTGGAG 1355  
QY 461 LysThrCysArgGlnValTyrGlnPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB AAACATGTAGACAGGTGTATGAGTTTATGATCAAGATCATGATCATGATCATGATCAT 1415  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysAspGlyHisArgLeuThrAla 500  
DB CCCCTGAGATGAGATACCTCTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1475  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB GCACATGCAAGAAATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB CAACCTGTATATTCACCGGAGCTTTGTACAGTTGTGTGCTGCTGCTGCTGCTGCTG 1595  
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
DB AATTTTGTGAAAGTTTGTCAATGTATGATCAAGTGTCAAAACCGCTTCCCGGAGTGC 1655  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValAlaArgLysCys 580  
DB CGCTGCAAGAGACAGTGCACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1715  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB GACCTGACCTCTGCTTACTTGTGAGCGGCTGACCATTTGAGCAAGTAAATATGTGCC 1775  
QY 601 CysLysAsnCysSerIleGlnArgLysSerLysLysHisLeuLeuLeuAlaProSerAsp 620  
DB TGCAAGAACTGACATATTCAGCGGAGCTCAAAAGCAATCTATATTCAGGAGCCATCTG 1835  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGln 640



QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleuGluSerIlyAspThrAspSer 380  
DB 1056 AACAGTAGAGGCCCCAGACCCCACTTAATGCTGGAAATCAAAGATCAAGACACT 1115  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyIuAsnAsnAspLyGluGluGluIlyAs 400  
DB 1116 GATTAGGGAAGCAGGAGCTGAACGGGGGAGAGAACATGATTAAGAGAGAGAGAGAG 1175  
QY 401 LyAspArgIuThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1176 AAAAGATGAACCTTCAGAGCTCCTCTGAGACCAATTCGCGTGTCAAAACCAATTAAGATG 1235  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1236 AAGCCAAATATTGAACCTCTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295  
QY 441 ArgValIleuIleGlyThrIleThrIleAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1296 AGAGTCTCATTTGGCAGCTTACTATGACAAATTTCTGTGCAATTCGTAGTTAATGGGACC 1355  
QY 461 LysThrCysArgGlnValIleGluPheArgValIleGluSerSerIleIleAlaProAla 480  
DB 1356 AAAACATGTAGACAGGTGTATGATTAGAGTCAAAAGATCTAGCATCATAGCTCAGCT 1415  
QY 481 ProAlaGluAspValAspThrProProArgIlyIlyIlyAspGlyIleIleArgLeuThrAla 500  
DB 1416 CCGGCTGAG 1475  
QY 501 AlaHisCysArgLyIleGlnLeuIlyIlyAspGlySerSerAsnHisValIleThrAsnIleTyr 520  
DB 1476 GCACTGTCAGAAAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1536 CAACCTGTGATCATTCAG 1595  
QY 541 AsnPheCysGluIlyPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1596 AATTTTGTGAAAGATTTTGTCAATGTAGTCAAGATGTCAAACCGCTTCCGGAGATGC 1655  
QY 561 ArgCysLyAspAlaGlnCysAsnThrIlyGlnCysProCysIleuAlaValArgIleCys 580  
DB 1656 CGCTGCAAGCAGCTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1715  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIleTrpAspSerIlyAsnValSer 600  
DB 1716 GACCTGACCTGTCTTACTGTGTGAGCGCTGACCAATGGGACAGTAAATGTGTCC 1775  
QY 601 CysLyAsnCysSerIleGlnArgGlySerIlyIlyIleuLeuLeuAlaProSerAsp 620  
DB 1776 TGCAGAACTGCAATTCAG 1835  
QY 621 ValAlaGlyIleTrpGlyIlePheIleIlyAspProValGlnIlyAsnGluPheIleSerGlu 640  
DB 1836 GTGGAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1895  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgIlyIlyValIleTrpAspLys 660  
DB 1896 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGAGAGAGAGAGAGAGAG 1955  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValIleAspAlaThrArgLys 680  
DB 1956 TACATGTGCAAGCTTCTGTCAACTGAAACAATGATTTGTGTGATGCAACCGCAGAG 2015  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaIlyValMet 700  
DB 2016 GGTAACAAATTCGTTTGCATTAATCATTCGTAATTCCTGCTATGCAAAAGTTATG 2075  
QY 701 MetValAsnGlyAspHisArgIleGlyIlePheAlaIlyAspArgAlaIleGlnThrGlyGlu 720  
DB 2076 ATGTTAACGGTATCAGAGATAGTATTTTGTCCAGAGAGAGAGAGAGAGAGAGAGAG 2135

QY 721 GluLeuPhePheAspTyrArgTyrSerGlnAlaAspAlaLeuIlyIleValGlyIleGlu 740  
DB 2136 GAGCTGTTTGTGATTACAGATTCAGCCAGGCTGATGCTGAAATGTCTGGCATTCGAA 2195  
QY 741 ArgGluMetGluIlePro 746  
DB 2196 AGAGAAATGAAATCCCT 2213

Search completed: August 13, 2006, 18:01:21  
Job time : 2236.12 secs



THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: August 13, 2006, 16:47:29 ; Search time 290.656 Seconds

(without alignments)  
6113.950 Million cell updates/sec

Title: US-10-773-302-2

Perfect score: 4051  
Sequence: 1 MGORGKSEKPEVCMRRKRVK.....RYSQADALKYGIEMREIRP 746

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2200221 seqs, 794037843 residues

Total number of hits satisfying chosen parameters: 4400442

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp  
-Q=/abs/ABSSWB\_epool/US10773302/runat\_11082006.140243.16928/app\_query.fasta.1  
-DB=Published Applications NA New -OPMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=D1osum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs03p  
-USER=US10773302 @CGN 1 1 430 @runat\_11082006.140243.16928 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA New:\*  
1: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
2: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
4: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
5: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
6: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
7: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
8: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
9: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
10: /BMC\_Celerra\_SIDS3/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4030.5	99.5	2253	US-10-539-630-2	Sequence 2, Appli
2	4030.5	99.5	2655	US-11-266-748A-25769	Sequence 25769, A
3	4030.5	99.5	2695	US-10-539-630-3	Sequence 3, Appli
4	4030.5	99.5	2695	US-11-266-748A-29501	Sequence 29501, A
5	2863	70.7	1803	US-11-266-748A-259286	Sequence 259286,
6	2863	70.7	1803	US-11-266-748A-278870	Sequence 278870,
7	2863	70.7	1803	US-11-266-748A-319803	Sequence 319803,

8	2603.5	64.3	4606	US-11-266-748A-62177	Sequence 62177, A
9	2603.5	64.3	4640	US-11-266-748A-26094	Sequence 26094, A
10	1817.5	44.9	2074	US-11-266-748A-186457	Sequence 186457, A
11	1817.5	44.9	2074	US-11-266-748A-241375	Sequence 241375, A
12	1785	44.1	1140	US-11-266-748A-98588	Sequence 98588, A
13	1785	44.1	1140	US-11-266-748A-151399	Sequence 151399, A
14	1536	37.9	1011	US-11-266-748A-175796	Sequence 175796, A
15	1416.5	35.0	904	US-11-266-748A-361325	Sequence 361325, A
16	1416.5	35.0	904	US-11-266-748A-444604	Sequence 444604, A
17	1369	33.8	1000	US-11-266-748A-283007	Sequence 283007, A
18	1369	33.8	1000	US-11-266-748A-209647	Sequence 209647, A
19	1350	33.3	1000	US-11-266-748A-292953	Sequence 292953, A
20	1350	33.3	1000	US-11-266-748A-244382	Sequence 244382, A
21	1350	33.3	1000	US-11-266-748A-404820	Sequence 404820, A
22	1350	33.3	1000	US-11-266-748A-475865	Sequence 475865, A
23	982	24.2	746	US-11-266-748A-357417	Sequence 357417, A
24	982	24.2	746	US-11-266-748A-440796	Sequence 440796, A
25	930	23.0	904	US-11-266-748A-216446	Sequence 216446, A
26	930	23.0	904	US-11-266-748A-238252	Sequence 238252, A
27	878	21.7	510	US-11-266-748A-8875	Sequence 8875, Ap
28	878	21.7	510	US-11-266-748A-63959	Sequence 63959, A
29	878	21.7	510	US-11-266-748A-66791	Sequence 66791, A
30	835.5	20.6	3366	US-11-218-305-17787	Sequence 17787, A
31	834	20.6	3133	US-10-953-349-35466	Sequence 35466, A
32	814	20.1	3050	US-10-953-349-37879	Sequence 37879, A
33	811.5	20.0	3194	US-11-218-305-17786	Sequence 17786, A
34	798	19.7	751	US-11-266-748A-74138	Sequence 74138, A
35	798	19.7	751	US-11-266-748A-108209	Sequence 108209, A
36	798	19.7	751	US-11-266-748A-126949	Sequence 126949, A
37	779.5	19.2	2512	US-11-216-545-4115	Sequence 4115, Ap
38	758.5	18.7	3401	US-11-218-305-17766	Sequence 17766, A
39	640.5	15.8	1250	US-11-216-545-6182	Sequence 6182, Ap
40	609	15.0	651	US-11-266-748A-355986	Sequence 355986, A
41	609	15.0	651	US-11-266-748A-439365	Sequence 439365, A
42	599.5	14.8	897	US-11-266-748A-190098	Sequence 190098, A
43	547	13.5	603	US-11-266-748A-14442	Sequence 14442, A
44	542	13.4	2313	US-11-266-748A-351035	Sequence 351035, A
45	542	13.4	2313	US-11-266-748A-383332	Sequence 383332, A

#### ALIGNMENTS

RESULT 1  
US-10-539-630-2  
; Sequence 2, Application US10539630  
; Publication No. US20060104981A1  
; GENERAL INFORMATION:  
; APPLICANT: Takeda Chemical Industries, Ltd.  
; TITLE OF INVENTION: Preventing and treating agent for cancer  
; FILE REFERENCE: 3130KOP  
; CURRENT APPLICATION NUMBER: US/10/539, 630  
; CURRENT FILING DATE: 2005-06-17  
; PRIOR APPLICATION NUMBER: JP2002-373144  
; PRIOR FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 14  
; SEQ ID NO 2  
; LENGTH: 2253  
; TYPE: DNA  
; ORGANISM: Human  
US-10-539-630-2

#### Alignment Scores:

Pred. No.: 0  
Score: 4030.50  
Percent Similarity: 99.2%  
Best Local Similarity: 99.5%  
Query Match: 6  
Gaps: 1

US-10-773-302-2 (1-746) x US-10-539-630-2 (1-2253)

Uy 1 MetGlyInrGlyLysLysSerGlyLysProValCysTrpArgValLys 20

Db 1 ATGGGCCAGACTGGGAAAGAAATCTGAGAGGGAACCAAGTTGTTGGCGGAGCGGTGTA 60  
Qy 21 SerGIuTyrMeCArgLeuArgGluLeuYsArgPheArgAlaAspGluValIysSer 40  
Db 61 TCAGAGTACATGCCAGCTGAGACAGCTCAAGAGGTTCCAGACGAGCTGATGAAGTAAAGCT 120  
Qy 41 MetPheSerSerAspAsnArgGlnIysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
Db 121 ATGTTTGGTTCCAATCCGTCAAGAAAATTTTGAAAAGAACGGAATCTTAAACCAAGAAATGG 180  
Qy 61 LyseGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 181 AAACGCGAAGAGATACAGCTGTGCAATCCGACTTGTAGGCTCAATCGCGGGAAT 240  
Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIysThr 100  
Db 241 AGGAGAGTTCGGTGCACAGTCACTTGATTTCCACACAGTCAATCCATTAAGACT 300  
Qy 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPhe 120  
Db 301 CTGATGCACTGGCTCAGTACCAATATGTAATCTTGCTCCCTACACAGCAATTTT 360  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
Db 361 ATGGTGGAGAGTAACTGTTTACATACATCTCTTATATGGAGATGAAGTTTATGAT 420  
Qy 141 GlnAspGlyThrPheIleGluGluLeuIleIysAsnTyrAspGlyIysValHisGlyAsp 160  
Db 421 CAGAGTGTACTTTTCATTGAAGAACTAATAAAAATTATGATGGGAAAGTACACGGGAT 480  
Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 481 AGAAGATGTGGGTTTATTAATGAAGAAATTTTGTGGAGTGTGGAAAGCCCTTGCTCAA 540  
Qy 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLeu 200  
Db 541 TATATATGATGATGAGATGATGATGATGAGACATCTGAAAGAAAGAAAGAAAGCAG 600  
Qy 201 LysAspLeuGluAspHisAsArgAspAspIysGluSerArgProProArgIysPheProSer 220  
Db 601 AAAAGTCTGGAGATCACCGAGATGATAAAGAAAGCCGCCACCTCGGAAATTTCTTCT 660  
Qy 221 AspIysIlePheGluAlaIleSerSerMetPheProAspIysGlyThrAlaGluGluLeu 240  
Db 661 GATAAATATTTTGAAGCCATTTCCATGTTTCCAGATTAAGGGGACGCGAAGAACTA 720  
Qy 241 LysGluIysTyrIysGluLeuThrGluGlnGluLeuProGluValAlaLeuProGluCys 260  
Db 721 AAGGAAAATATTAAGAACTCACGAAACGAGCTCCAGGGGCACTTCTCTGAAATGT 780  
Qy 261 ThrProAsnIleAspGlyProAsnAlaIysSerValGlnArgGluGlnSerLeuHisSer 280  
Db 781 ACCCCCAACATAGATGAGCAAAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACACTCC 840  
Qy 281 PheHisThrLeuPheCysArgArgCysPheIysTyrAspCysPheLeuHis----- 297  
Db 841 TTTCAATCGCTTTTCTGTAGCGAGTGTAAATGATGACTCTTCTCATCTGTAAGTGC 900  
Qy 298 -----ProPheHisAlaThrProAsnThrTyrIysArgIysAsnThrGluThrAlaLeu 315  
Db 901 AATTATTTCTTTTCAGCAACACCCAAACCTTATTAAGCCGAAAGAACAGAAACAGCTTA 960  
Qy 316 AspAsnIysProCysGlyProGluCysTyrGlnHisLeuGluGluValAlaIysGluPheAla 335  
Db 961 GACAAACAACCTTGTGAGCAACAGTGTACAGCAATTTGAGGGAGCAAAAGAGTTTGTCT 1020  
Qy 336 AlaAlaLeuThrAlaGluArgIleIysThrProProIysArgProGlyIysArgArgArg 355  
Db 1021 GCTGCTCTCAACCGCTGACGGATTAAGACCCCAACAAACGTCCAGAGGGCGCGAAGGA 1080  
Qy 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGlnSer 375  
Db 1081 GAGCGGCTTCCCAATTAACATGACAGGCCCAACACCCCAACATTAATGCTGGAATCA 1140

Qy 376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyIleAsnAsnAspIys 395  
Db 1141 AAGGATACAGACAGTGTATAGAGAAACGAGCACTGMAACGGGGGAGGAACAATGATTA 1200  
Qy 396 GlnGluGluGluIysIysAspGluThrSerSerSerSerGlnAlaAsnSerArgCysGln 415  
Db 1201 GAAGAGAGAGAAAGAAAGATGAATCTCAAGCTCTCTGAAGCAAAATTTCTCGGTGTCAA 1260  
Qy 416 ThrProIleIysMetCysProAsnIleGluProProGluAsnValGluTrpSerGlyAla 435  
Db 1261 AACACATTAAGATGAAGCCAAATATGAACCTCTGAGATGTGAGTGAAGTGTGCT 1320  
Qy 436 GlnAlaSerMetPheArgValLeuIleGlyThrTyrTyrAsnAsnPheCysAlaIleAla 455  
Db 1321 GAAGCTCAATGTTTGAAGTCTCATATGGCACTTACATGACAATTTCTGTGCAATGCT 1380  
Qy 456 ArgLeuIleGlyThrIysThrCysArgGlnValTyrGluPheArgValIysGluSerSer 475  
Db 1381 AGGTTAATGGAGCAAAACATGTAGACAGGTGTATGAGTTTATGATGTCMAAGAACTTAC 1440  
Qy 476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgIysIysValIys 495  
Db 1441 ATCATAGCTCAGCTCCGCGTGAAGATGGATGTGATCTCTCCAAAGAAAAGAAAGAGAA 1500  
Qy 496 HisArgLeuThrAlaAlaHisCysArgIysIleGlnLeuIysIysAspGlySerSerAsn 515  
Db 1501 CACCGTGTGGGTGACACTGCAAGAAAGTACAGCTGAAGAAAGACGCGCTCTTCTTAC 1560  
Qy 516 HisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysPro 535  
Db 1561 CATGTTTACATATCAACCTGTGATCATTCACGAGCCTTGTGACAGTTCTGTGCTCT 1620  
Qy 536 CysValIleAlaGlnAsnPheCysGluIysPheCysGlnCysSerSerGluCysGlnAsn 555  
Db 1621 TGTGTATAGCACAAAATTTTGTGAAAGTTTGTCAATGTAGTTAGAGTCAAGTCAAAAC 1680  
Qy 556 ArgPheProGlyCysArgCysIysValGlnCysAsnThrIysGlnCysProCysTyrLeu 575  
Db 1681 CGCTTTCGGGATGCCCTGCAAGCAACGTGCAACCAAGCTGCCGTGCTTACTGTG 1740  
Qy 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIleTrpAsp 595  
Db 1741 GCTGTCCAGAGTGTGACCTGACCTGTCTTACTGTGTGAGCCGCTGACCATTTGGGAC 1800  
Qy 596 SerIysAsnValSerCysIysAsnCysSerIleGlnArgGlySerIysIysIleLeu 615  
Db 1801 AGTAAATATGTCTCTCAAGAACTGCAGATTTCAAGCGGGGCTCCAAAAGCATCTATTG 1860  
Qy 616 LeuAlaProSerAspValAlaGlyTyrGlyIlePheIleIysAspProValGlnIysAsn 635  
Db 1861 CTGGCAACATCTGACGTGGGAGGCTGGGGGATTTTATCAAAAGTCTGTGTCMAAAAT 1920  
Qy 636 GlnPheIleSerGluTyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGly 655  
Db 1921 GAATTCATCTCAGAAATCTGTGAGAAATTTCTCAAGATGAAGCTGACAGAAAGGG 1980  
Qy 656 LysValTyrAspIysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheValVal 675  
Db 1981 AAGTGTATGATTAATACATGTGCAAGCTTCTGTTCAACTTGAACATGATTTGTGTGTG 2040  
Qy 676 AspAlaThrArgIysGlyAsnIysIleArgPheAlaAsnHisSerValAsnProAsnCys 695  
Db 2041 GATCAACACCGCAAGGGTGAACAAATTCGTTTGCAATCATTTGCGTAAATCCAAATGTC 2100  
Qy 696 TyrAlaIysValIleMetMetValAsnGlyAspHisArgIleGlyIlePheAlaIysArgAla 715  
Db 2101 TATCAAAAGTTAATGATGATTAACGGTGTATCAAGATAGATATTTTGCMAAGAGCC 2160  
Qy 716 IleGlnThrGlyGluGluLeuPheAspTyrArgTyrSerGlnAlaAspAlaLeuIys 735  
Db 2161 ATCCAGACTGGCGAAGAGCTGTTTGTGATTACAGATACAGCCAGGCTGATGCCCTGAG 2220

QY 736 TTTTGGGATCGGAAAGAAATGAAATCCCT 2253  
Db 2221 TATGTCGATCGGAAAGAAATGAAATCCCT 2253  
RESULT 2  
US-11-266-748A-25769  
; Sequence 25769, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harklin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 25769  
; LENGTH: 2655  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-11-266-748A-25769  
Alignment Scores:  
Pred. No.: 0 Length: 2655  
Score: 4030.50 Matches: 745  
Percent Similarity: 99.2% Conservative: 0  
Best Local Similarity: 99.2% Mismatches: 1  
Query Match: 99.5% Indels: 5  
DB: 8 Gaps: 1  
US-10-773-302-2 (1-746) x US-11-266-748A-25769 (1-2655)  
QY 1 MetGlyGlnThrGlyLeuLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
Db 123 ATGGCCAGACTGGGAAAGAAATCTGAAAGGAGCAAGTTGTGGCGGAAGCGTGTAAAA 182  
QY 21 SerGlyLysMetArgLeuArgGlnLeuLysArgPheArgAlaAspGlnValLysSer 40  
Db 183 TCAGAGTCAATCGACTGAGACAGCTCAAGAGTTCAAGCACTGATGAATGAAGAGT 242  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGlnIleLeuAsnGlnGlnLys 60  
Db 243 ATGTTTATGTTCCAAATCGTCAGAAATTTTGAAAGAACGGAATCTTAAACCAAGAAAGG 302  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgLysThr 80  
Db 303 AAACGAGCAAGATACAGCTGTGCACTGACTTCTGTAGCTCATTTGCGCGGAGCT 362  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 363 AGGAGGTGTTGGTGAACAGTACTGATTTTCCAAACCAAGTATCCCATTAAGACT 422  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnGlnAsnPhe 120

Db 423 CTGAATGCAAGTTGCTTACAGTACCATAATGATTTCTGGTCTCCCTCAACAGCAATTTT 482  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProLysMetGlyAspGlnValLeuAsp 140  
Db 483 ATGTGTGAAGATGAAGAACTGTTTACATTAACATTTCTTATATGAGGAGATGAAGTTTGAAT 542  
QY 141 GlnAspGlyThrPheIleGluLysLeuLysAsnIleLysAsnIleLysAsnIleLysAsn 160  
Db 543 CAGATGTGATCTTCAATTAAGAACTAATTAATAATTTATGATGGAAAGTACACGGAGAT 602  
QY 161 ArgGluCysGlyPheIleAsnAspGlnIlePheValGluLeuValAsnAlaLeuGln 180  
Db 603 AGAGATGTGGGTTATTAATATGATAAATTTTGGAGTTGGTGAATGCCCTTGGTCA 662  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgLysGln 200  
Db 663 TATATATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722  
QY 201 LysAspLeuGluAspHisArgAspAspLysGlnSerArgProArgLysPheProSer 220  
Db 723 AAGATCTGAGATCAACGAGATATTAAGAAAGCCGCCCTCGAAATTTCTTCT 782  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluLysLeu 240  
Db 783 GATTAATTTTGAAGCCATTTCTCAATGTTTCAATATGAGGCAACAGCAAGAACTA 842  
QY 241 LysGluLysLysLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProGluLys 260  
Db 843 AAGGAAATATTAAGAACTACCAACAGCAAGCTCCAGGCGCACTTCTCTTAATGT 902  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnIleSerLeuHisSer 280  
Db 903 ACCCCCAATATGATGACCAATCTTAATCTTCAAGAGACCAAACTTACACTCC 962  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysLysAspCysPheLeuHisS----- 297  
Db 963 TTTGATACGCTTTTCTGATAGGAGATGTTTAAATATGATGATGATGATGATGATGATGATGAT 1022  
QY 298 -----ProPheHisAlaThrProAsnThrTyrLysAspGlyAsnThrGluThrAlaLeu 315  
Db 1023 AATTATTTCTTTTCAGCAACACCCCAACCTTATTAAGCGAAGAACAGCAACACCTCTA 1082  
QY 316 AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluGluValAlaLysGluPheAla 335  
Db 1083 GACACCAATCTTGTGACCAACAGTGTACAGATTTGAGAGGAGCAAAAGATTTGCT 1142  
QY 336 AlaAlaLeuThrAlaGluArgIleLysThrProProLysArgProGlyValArgArgArg 355  
Db 1143 GCTGCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTCCAGAGGCGCGCAAGAA 1202  
QY 356 GluArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375  
Db 1203 GACCGCTTCCCAATTAACATGAGCAGGCGCCAGACCCCAACATTAATGTGTGAACA 1262  
QY 376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGluValAsnAsnAspLys 395  
Db 1263 AAGATACAGACAGGATAGGAG 1322  
QY 396 GluGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415  
Db 1323 GAAAGAAAG 1382  
QY 416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAla 435  
Db 1383 ACACCAATTAAGATGAAGCAAAATTAATGAACCTCTGAGAAATGAGATGAGATGAGTCT 1442  
QY 436 GluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla 455  
Db 1443 GAAAGCTTAATGTTTGAAGTCTCAATGCGCTTACTATGACAAATTTCTGAGCAATTTCT 1502  
QY 456 ArgLeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer 475  
Db 1503 AGGTTAATTTGAGCAAAACATGTATGACAGGTATGATGATTTAAGTCAAAAGATCTAGC 1562

QY 476 IleIleAlaProAlaProAlaGluAspValAspThrProAlaGlyValLeuAspArgLys 495  
DB 1563 ATCATAGCTCCAGCTCCGCTGAGGATGTGATACTCTCCAGAGAAAAGAGAGGAAA 1622  
QY 496 HisArgLeuThrPalaHisCysArgGlyLeuGluLeuLeuValAspGlySerSerAsn 515  
DB 1623 CACCGTTGTGGCTGCACACTGCAGAAAGATACAGCTGAAAAGAGAGCGCTCTAC 1682  
QY 516 HisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysPro 535  
DB 1683 CATGTTTACACTATACACCTGTGATCATCCAGCGACCTGTGACAGTTCCGTCCTT 1742  
QY 536 CysValIleAlaGlnAsnAspCysGluValAspCysGlnCysSerSerGluCysGlnAsn 555  
DB 1743 TGTGTGATAGCAAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGATGTCCAAAC 1802  
QY 556 ArgPheProGlyCysArgCysLeuAlaGlnCysAsnThrLysGlnCysProCysThrLeu 575  
DB 1803 CGCTTTCGGGATGCCCTGCAGAGCAGATGCACACAGCAGTCCCTGTCTACCTG 1862  
QY 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisThrAsp 595  
DB 1863 GCTGTCCAGAGTGTGACCTGACCTCTGTCTTACTTGTGAGCGCTGACCATTTGGAC 1922  
QY 596 SerLysAsnValSerCysLeuAsnCysSerIleGlnArgGlySerLysLeuLeu 615  
DB 1923 AGTAAAAATGTGCTCGCAGAAAGCTGCAGTATTCAGCGGGCTCCAAAACATCTATTG 1982  
QY 616 LeuAlaProSerAspValAlaGlyThrGlyLeuPheIleLysAspProValGlnLysAsn 635  
DB 1983 CTGGACCATGTGACGTGCGAGGCTGGGGGATTTTATCAAGATCCGTGTCAAGAAAAT 2042  
QY 636 GluPheIleSerGluTyrCysGlyGluLeuIleSerGlnAspGluAlaAspArgArgGly 655  
DB 2043 GAATTCATCTCAGAAATCTGTGAGAGATTTTCTCAAGATGAAGCTGACAGAGAGGG 2102  
QY 656 LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAla 675  
DB 2103 AAAGGTATGATTAATATACATGTGCGCTTCTGTTCACTTGAACTTATTTGTGTGTG 2162  
QY 676 AspAlaThrArgLysGlyValAsnLysIleArgPheAlaAsnHisSerValAsnProAsnAsp 695  
DB 2163 GATGCAACCCGCAAGAGGTACAAATTCGTTTGCAAATTCATTCGTAATCCAAACTGC 2222  
QY 696 TyrAlaLysValMetMetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAla 715  
DB 2223 TATCAGAAAGTTATGATGTGTTAACGCTGATCAGAGATAGGTAATTTTGGCAAGAGACC 2282  
QY 716 IleGlnThrGlyGluGluLeuPheAspTyrArgTyrSerGlnAlaAspAlaLeuLys 735  
DB 2283 ATCCAGACTGGCGAAGAGCTGTTTGTATTAAGATACAGCCAGCGCTGATGCCCTGAAG 2342  
QY 736 TyrValGlyIleGluArgGluMetGluIlePro 746  
DB 2343 TATGTCCGATCGAAAGAGAAATGAAATCCCT 2375

! ORGANISM: Human  
US-10-539-630-3  
Alignment Scores:  
Pred. No.: 0  
Score: 4030.50  
Percent Similarity: 99.2%  
Best Local Similarity: 99.2%  
Query Match: 99.5%  
DB: 6  
Gaps: 1  
US-10-773-302-2 (1-746) x US-10-539-630-3 (1-2695)  
QY 1 MetGlyGlnThrGlyValLysSerGluLysGlyProValCysThrArgLysArgValLys 20  
DB 167 ATGGGCGACACTGGGAGAAATCGAAGAGGACCAATTTGTGGCGGAGCGGTAAAC 226  
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 227 TCAGAGTACATGCCACTGACAGCTCAGAGAGGTTTCAGACGAGCTGATGAAGTAAAGAT 286  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnLysTrp 60  
DB 287 ATGTTTATGTTCCATTCGTCAAGAAATTTGGAAGAACGGAATCTTAAACCAAGAAATGG 346  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACAGGAGAGATACAGCTGTGCACATCTCTGACCTCATTTGCCGGGACT 406  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 407 AGGAGAGTGTCCGTGACCAAGTACTTGGAATTTTCCACACAGTCAATCCATTAACACT 466  
QY 101 LeuAlaAlaValAserValProIleMetTyrSerTrpSerProLeuGlnAsnAsp 120  
DB 467 CTGAATCAGTGTCTTCAATGATCCATATGATATCTTGATCCCTACAGAGAAATTTT 526  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 527 ATGTGTGAAGATGAACTGTTTATCAATACATTCCTTATATGGAGATGAAGTTTGTGAT 586  
QY 141 GluAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 587 CAGATGTACTTTCAATTGAAGAACTAATTAATAATTAATGATGGAGAAAGTACAGGGGAT 646  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAlaAlaLeuGlyGln 180  
DB 647 AGAAGATGTGGTTTATTAATGATGAATTTTGTGAGTTGGGAATGCCCTTGGTCAA 706  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluGln 200  
DB 707 TATATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
QY 201 LysAspLeuGluAlaAspHisArgAspAspLysGluSerArgProAlaGlyLysPheProSer 220  
DB 767 AAAGATCTCGAGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCCGAAATTTCTTCT 826  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 827 GATTAATTTTGTGAAGCATTTCTCTCAATGTTTCAAGTAAAGGCAACAGCAAGAGAACTA 886  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGluCys 260  
DB 887 AAGGAAAAATTAAGAATCCACAGCAAGAGAGTCCAGGCGCATCTCCCTCAAGATGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 947 ACCCCCAATGATGAGCAACAAATGCTAAATCTGTCAAGAGAGCAAAAGCTTACACTCC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHis----- 297  
DB 1007 TTTGATACGCTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTTCAATCGTAAGTGC 1066

QY 298 -----ProPheH1s1aThrProAsnThrTyrLysArgLysAsnThrGluThra1a1eun 315  
 DB 1067 AATATTCCTTTCAGACACCCCACTTATTAAGCGAAGAACACAGAAACAGCTCTA 1126  
 QY 316 AspaenLysProCysGlyProGlnCysTyrGlnH1sLeuGluGlyAlaLysGluPheAla 335  
 DB 1127 GACACCAAACTTGTGAGACACAGTGTTCACAGACTTGTGAGAGCAAAAGAGTTGTCT 1186  
 QY 336 AlaAlaLeuThra1aGluArgL1sLeuThrProProLysArgProGlyGlyLysArgArg 355  
 DB 1187 GCTGCTTCACCGCGAGCGAGATTAAGACCCCAACAAACGTCAGAGAGCGCGAGAA 1246  
 QY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThr11AsnValLeuGlnSer 375  
 DB 1247 GGAAGGCTTCCCAATACAGTACAGAGCCACAGACCCCACTTAATATGTCTGAAATCA 1306  
 QY 376 LysAspThraAspSerAspArgL1a1aGlyThrGluThrGlyGlyLysAsnAsnLys 395  
 DB 1307 AAGATACAGACAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 1366  
 QY 396 GluGluGluGluLysLysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGln 415  
 DB 1367 GAAG 1426  
 QY 416 ThrPro11sLeuMetLysProAsn11eGluProProGluAsnValGluThrSerGlyAla 435  
 DB 1427 ACACCAATTAAGATGAGAGCAATATTAACCTTCGAGAAATGAGAGAGAGAGAGAG 1486  
 QY 436 GluAlaSerMetPheArgValLeu11eGlyThrTyrTyrAspAsnPheCysAla11eAla 455  
 DB 1487 GAAGCTCAATGTTTATGAGTCTCTCATTTGACCTTAAGCAATTTCTGCTCCATTGCT 1546  
 QY 456 ArgLeu11eGlyThrLysThrCysArgGluValTyrGluPheArgVal11sGlnSerSer 475  
 DB 1547 AGGTAAATTTGGAGCAAAACATGATGACAGGTATGAGTTATGAGTCAAAAGATCTAGC 1606  
 QY 476 11e11eAlaProAlaProAlaGluAspValAspThrProProArgLysLysArgLys 495  
 DB 1607 ATCATAGCTCCAGCTCCGCTAGAGATGAGATCTCCCAAGAGAAAGAGAGAGAGAA 1666  
 QY 496 HisArgLeuThra1a1aH1sCysArgLys11eGlnLeuLysLysAspGlySerSerAsn 515  
 DB 1667 CACCGGTTGTGGGCTGACACACAGAGAGAGATACAGCTGAAGAGAGAGAGAGAGAG 1726  
 QY 516 HisValTyrAsnTyrGlnProCysAspH1sProArgGlnProCysAspSerSerCysPro 535  
 DB 1727 CATGTTTCAACTATCAACCCCTGATCATCCACGCGAGCTTGTGACAGTTCTGCGCT 1786  
 QY 536 CysVal11eAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsn 555  
 DB 1787 TGTGTGATGACCAAAATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTCAAAAC 1846  
 QY 556 ArgPheProGlyCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeu 575  
 DB 1847 CGCTTCCGAGATGCGCTGCAACACAGTCAACACCAAGCAATGCGCGCTGACCTG 1906  
 QY 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspH1sTrpAsp 595  
 DB 1907 GCTGTCCAGAGATGACCTGACCTGTCTTACTTGTGAGCGCGCTGACCATTTGGAGC 1966  
 QY 596 SerLysAsnVal11sCysLysAsnCysSer11eGlnArgGlySerLysLysLysLeu 615  
 DB 1967 AGTAAATATGTCTCTGCAAGAACCTGATATTCAGCGGCTTCAAAAAGATCTATTG 2026  
 QY 616 LeuAlaProSerAspValAlaGlyTyrGly11ePhe11eLysAspProValGlnLysAsn 635  
 DB 2027 CTGGACCATTCAGATGAGTGGAGAGCTGGGGAATTTTATCAAAAGATCTCTGACAGAAAT 2086  
 QY 636 GluPhe11sSerGlyTyrCysGlyGlu11e11sSerGlnAspGluAlaAspArgArgGly 655  
 DB 2087 GAATTCATCTCAGATTAAGTATGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAGAG 2146  
 QY 656 LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValVal 675

DB 2147 AAGGTATGATTAATATCATGTCAGCTTCTGTTCACTTGAACATGATTTTGGGTG 2206  
 QY 676 Aspa1aThraGlyGlyLysLys11eArgPheAlaAsnH1sSerVal1AsnProAsnCys 695  
 DB 2207 GATGCAACCCCGCAAGGATTAACAAATTCGTTTGTCAAAATTCATTCGTTAAATCAACAGTC 2266  
 QY 696 TyrAlaLysVal11MetMetValAsnGlyAspH1sArg11eGly11ePheAlaLysArgAla 715  
 DB 2267 TATGCAAAAGTTATATGATGATTAACGATATCAAGAGATGATTTTGTCCAAAGAGAGCC 2326  
 QY 716 11eGlnThrGlyGluGluLeuPheAspTyrArgTyrSerGlnAlaAspa1aLeuLys 735  
 DB 2327 ATCCAGACTGGCGAAGAGACTGTTTTCATTTACAGATACAGTCAGAGCTGATGTCCTGAAG 2386  
 QY 736 TyrValGly11eGluArgGluMetGlu11ePro 746  
 DB 2387 TATGCGGCATCGAAAGAAATGGAATCCCT 2419  
 RESULT 4  
 US-11-266-748A-29501  
 ; Sequence 29501, Application US/11266748A  
 ; Publication No. US20060134663A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harkin, Paul  
 ; APPLICANT: Johnston, Patrick  
 ; APPLICANT: Mulligan, Karl  
 ; TITLE OF INVENTION: Transcriptome Microarray Technology and  
 ; FILE REFERENCE: 55815-0102 (319189)  
 ; CURRENT APPLICATION NUMBER: US/11/266,748A  
 ; CURRENT FILING DATE: 2005-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105479.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105482.6  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105483.4  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105507.0  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105485.9  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105484.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: US 60/662,276  
 ; PRIOR FILING DATE: 2005-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/700,293  
 ; PRIOR FILING DATE: 2005-07-18  
 ; NUMBER OF SEQ ID NOS: 483996  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 29501  
 ; LENGTH: 2695  
 ; TYP: DNA  
 ; ORGANISM: Homo Sapiens  
 US-11-266-748A-29501  
 Alignment Scores:  
 Pred. No.: 0 Length: 2695  
 Score: 4030.50 Matches: 745  
 Percent Similarity: 99.2% Conservative: 0  
 Best Local Similarity: 99.2% Mismatches: 1  
 Query Match: 99.5% Indels: 5  
 DB: 8 Gaps: 1  
 US-10-773-302-2 (1-746) x US-11-266-748A-29501 (1-2695)  
 QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
 DB 167 ATGGCCAGACTGGAGAGAAATCTGAGAGGAGACCAAGTTTGTGCGGAGACGCTATAAA 226  
 QY 21 SerGluTyrMetCysLeuArgGlnLeuLysArgPheArgGAlaAspGluValLysSer 40  
 DB 227 TCAAGATACATCGCACTGACAGCTCAAGAGGTTTCAAGACAGCTGATGAAGTAAAGAGT 286

QY 41 MetPheSerSerAenArgInLysIleuEngluArgThrGluIleuAenGlnGluTrp 60  
DB 287 ATGTTAGTTCCATTCGTCAAGAAATTTTGAAAGAACGAAATCTTAAACCAAGAAATGG 346  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACAGCGAAGAGATACAGCCTGTGCACATCTGTGACCTCTGACCTCAATTCGGGGAGCT 406  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 407 AGGAGGTGTTCCGTGACCAAGTACCTTGATTTTCCACACAAAGTACCTCCATTAAAGACT 466  
QY 101 LeuAenAlaValAlaSerValProIleMetTySerTrpSerProLeuGlnIleAspPhe 120  
DB 467 CTGATGACAGTTGGCTTCAGTACCAATATGATTTCTTGCTCTCCCTCAACAGAAATTTT 526  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyMetGlyAspGluValLeuAsp 140  
DB 527 ATGTGTGAAAGATGAACCTGTTTACATTAACATCTCTTATATGGAGATGAAGTTTATGAT 586  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyArgGlyLysValHisGlyAsp 160  
DB 587 CAGATGTACTTTCATTGAAGAACTAATTAATAATTTGATGGGAAAGTACACGGGAT 646  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 647 AGAGATGTGGGTTATTAATGAATGAATTTTGTGGAGTTGGTGAATGACCTTGTCATA 706  
QY 181 TyTrpAspAspAspAspAspAspAspGlyAspAspProGluGluValGluGln 200  
DB 707 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 766  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 767 AAAGTCTGGAGATCACCGAGATGATTAAGAAAGCCGCCACTCGGAAATTTCTCTT 826  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 827 GATTAATAATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGAACTA 886  
QY 241 LysGluLysGlyTyLysGluLeuThrGluGlnGluLeuProGlyAlaLeuProProGluLys 260  
DB 887 AAGAAAAATATTAAGAACTACACCGAAGACGCTCCAGCGCCTTCCTCGAAGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 947 ACCCCCAATATGATGACCAATATCTTAATCTGTCAGAGAGCAAGCTTACATCTCC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyArgCysPheLeuHis----- 297  
DB 1007 TTTTCAATAGCTTTTCTGTAGCGAATGTTTAAATATGATGATCTCTTCAATCGTAAGTGC 1066  
QY 298 -----ProPheHisAlaThrProAsnThrTyLysArgGlyAsnThrGluThrAlaLeu 315  
DB 1067 AATTATCTTTCATGCAACACCCCAACTATATAGCGAAGAACACAGAAACAGCTCTTA 1126  
QY 316 AspAsnLysProCysGlyProGlnCysTyArgHisLeuGluGlyAlaLysGluPheAla 335  
DB 1127 GACAAACAACCTTTGGACCAAGTGTATCCAGCATTTGGAGGGAGCAAGAGATTTGCT 1186  
QY 336 AlaAlaLeuThrAlaGluArgIleLysThrProProLysArgProGlyGlyArgArgArg 355  
DB 1187 GCTGCTCTCACCGCTGACGGATTAAGACCCCAAAACGTCCAGAGCGCGCAGAGA 1246  
QY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375  
DB 1247 GGAACGGCTTCCCAATACAGTACGAGCGCCACAGACCCCACTTAATGTCTGGAACA 1306  
QY 376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLys 395  
DB 1307 AAGGATACAGACGATGATGAGGAGCAAGGACTGAACGGGGGAGAGAAACATATATAA 1366

QY 396 GlnGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415  
DB 1367 GAAGAAAGAAAGAAAGAAAGTGAACCTTCAGACTCTCTAAGCAAAATTCCTGGTGTCAA 1426  
QY 416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAla 435  
DB 1427 ACACCAATTAAGATGAAGCCAAATATTGAACCTCTCGAAGATGTGGATGGAGTGGCT 1486  
QY 436 GlnAlaSerMetPheArgValLeuIleGlyThrTyTrpAspAspPheCysAlaIleAla 455  
DB 1487 GAAGCTCAATGTTAAGTTCATCTCAATGGCATTAATGATCAATTTCTGTGCATTTGCT 1546  
QY 456 ArgLeuIleGlyThrLysThrCysArgGlnValTyArgLysPheArgValLysGluSerSer 475  
DB 1547 AGGTAAATGGGACCAAAACATGACAGGTGTATAGTTTGAATGAATCAAAATTTAC 1606  
QY 476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysValArgLys 495  
DB 1607 ATCATAGCTCAGCTCCCGCTGAGATGTGATCTCTCCAGAGAAAGAAAGAGAA 1666  
QY 496 HisArgLeuTrpAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsn 515  
DB 1667 CACCGGTGTGGGCTGCACCTGCAGAAAGATACAGCTGAAGAAAGACGGCTCTCTTAAC 1726  
QY 516 HisValTyTrpAsnTyArgInProCysAspHisProArgGlnProCysAspSerSerCysPro 535  
DB 1727 CATGTTCACATATCAACCTGTGATCATCCAGGAGCCCTGTGACAGTGTGAGCCCT 1786  
QY 536 CysValIleAlaGlnAspPheCysGluLysPheCysGlnCysSerSerSerGluCysGlnAsn 555  
DB 1787 TGTGTGATGACCAAAATTTTGTGAAGATTTTGTCAATGATGTTGATGATGATGATGAT 1846  
QY 556 ArgPheProGluCysArgCysLysValGlnCysAsnThrLysGlnCysProCysTyLeu 575  
DB 1847 CGCTTCCGGGATCCGCTCGAAGACAGATGCAACCAAGCAGTCCCGCTGCTACCTG 1906  
QY 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAsp 595  
DB 1907 GCTGTCCGAGGTGTGACCTGACCTGTCTTAATCTGTGAAGCCGTGACCATTTGGAC 1966  
QY 596 SerLysAsnValSerCysLysAsnCysSerIleGlnArgLysSerLysValLeuLeu 615  
DB 1967 AGTAAATATGTCTCTCAAGAACTGACAGTATTCAGCGGGGCTCCAAAGAACTATTTG 2026  
QY 616 LeuAlaProSerAspValAlaGlyTyArgIlePheIleLysAspProValGlnLysAsn 635  
DB 2027 CTGGCACCATCTGACGTGGCAGGCTGGGGAATTTTATCAAAAGATCTGTGCAGAAAAAT 2086  
QY 636 GlnPheIleSerGluTyCysGlyGlnIleIleSerGlnAspGluAlaAspArgArgGly 655  
DB 2087 GAATTCATCTCAGAAATCTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGG 2146  
QY 656 LysValTyArgLysTyMetCysSerPheLeuPheAsnLeuAsnAspPheValAla 675  
DB 2147 AAAGTGTATGATTAATACATGTGACGCTTTCTGTCAACTTGAACATATGATTTGTG 2206  
QY 676 AspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695  
DB 2207 GATCAACCCGCAAGGGTAAACAAATTCGTTTTCGAATCATTTGGTAAATCCAAATGCTC 2266  
QY 696 TyTrpAlaLysValMetMetValAsnGlyAspHisArgIleGlyIlePheAlaLysArgAla 715  
DB 2267 TATCAAAAGTTAAGATGTTAAGAGGTATCAAGATAGATATTTTGGCCAAAGAGCC 2326  
QY 716 IleGlnThrGlyGluGluLeuPhePheAspTyArgTySerGlnAlaAspAlaLeuLys 735  
DB 2327 ATCCAGATCGCGAAGAGCTGTTTGTGATTAACGATACAGCCAGGCTGATGCCCTGAAG 2386  
QY 736 TyTrpAlaGlyIleGluArgLysMetGlnIlePro 746  
DB 2387 TATGTCCGCAATCGAAAGAAATGAATCCCT 2419

RESULT 5





QY 510 SASPGlySerSerAsnHisValTyrAsnTyrGlnProCysAsp-HisProArgGlnProC 530  
DB 1342 GGACGGCTCTCTTAACCATGTTTCAACTCAACCCGTGATTCATCAAGCAGCCTT 1401  
QY 530 YSASPserSerCysProCysValIlealaglnAsnPhcCys-GluYsrPhcCysGlnCys 549  
DB 1402 GTGACAGTCTGTCCTTGTGTGTATGACAAATTTTGTGTAAGTTTGTCAATGT 1461  
QY 550 SerSer-GluCysGlnAsnArg-PheProGlyCysArgCysValalaglnCysAsnThrL 569  
DB 1462 AGTTCAGAGTGTCAAAACCCCTTCCGGGATCCGCTGCAAGACACAGTCAACCA 1521  
QY 569 YSGlnCysProCysTyr-LeuAlaValArgGluCysAspProAspLeuCysLeuThrCys 588  
DB 1522 AGCAGTGGCCGTGTACCTCGCTGTCCGAGAGTGTGACCTGTGCTTACTTGT 1581  
QY 589 GLYAlaAlaAspHisTyrPheSerSerLeuAsnValSerCysValAsnCysSerIleGlnArg 608  
DB 1582 GGACCCGCTGACCAATGGGACAGTAAATGTCTCGCAAGAACTGACGATTCAGCGG 1641  
QY 609 GlyseryslyshsleuLeuLeuAlaProSerAspVal-AlaGlyTyrGlyIlePheI 628  
DB 1442 GGCTCCAAAAGCATCTATTGCTGACCATCTGACGTCGCAAGCTGGGGATTTTAT 1701  
QY 628 eLYsAspProValGlnYsAsnGluPhe-IleSerGlyTyrCys-GlyGluIleIle-- 646  
DB 1702 CAAGATCTCTGACGAAATAATGAATTCATCTTCAAGATCTGTGGAGAGATTAATTT 1761  
QY 647 ----SerGlnAspGluAlaAspArgArgGlyValTyr 658  
DB 1762 CTNCAAGGTGAGAGGTGACAGATGAGGAGGAAAGTTTAT 1801

## RESULT 6

US-11-266-748A-278870  
Sequence 278870, Application US/11266748A  
Publication No. US2006013463A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OR INVENTION: Transcription Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIORITY FILING DATE: 2005-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2004-11-03  
PRIORITY FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: Patent version 3.3  
SEQ ID NO 278870  
LENGTH: 1803  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (356)-(379)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (1680)..(1680)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1764)..(1764)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1792)..(1792)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-278870

Alignment Scores:  
Pred. No.: 8,72e-233 Length: 1803  
Score: 2863.00 Matches: 549  
Percent Similarity: 94.8% Conservative: 1  
Best Local Similarity: 94.7% Mismatches: 18  
Query Match: 70.7% Indels: 12  
Gaps: 1

US-10-773-302-2 (1-746) x US-11-266-748A-278870 (1-1803)

QY 91 PheProThrGlnValIleProLeuYsrThrLeuAsnAlaValAlaSerValProIleMet 110  
DB 82 TTCCCGACAGAGTATCCCTTAATTAAGACTCTGAATCAATGCTTCACTAATGATGATG 141  
QY 111 TysSerTyrSerProLeuGlnGlnAsnPhcValGluAspGlyThrPheIleGluGluIle 130  
DB 142 TATCTTGTCCTCCCTTACAGCAAGATTTTATGTTGAGATGAACTGTTTATCATATAC 201  
QY 131 IleProTyrMetGlyAspGlyValIleAspGlnAspGlyThrPheIleGluGluIle 150  
DB 202 ATTCTTATATGAGAGATGAAATTTTATGATCAGATGTCCTTCACTGAAGAACTAATA 261  
QY 151 YsAsnTyrAspGlyValIleGlyAspArgGluCysGlyPheIleAsnAspGluIle 170  
DB 262 AAAAATTAATGATGGAAGATGACAGGGATGAGAAATCTGGCTTTAATGATGAATTT 321  
QY 171 PheValGluLeuValAsnAlaLeuGlyGlnTyrAsnAspAspAspAspAspGly 190  
DB 322 TTGTGAGGTGGTGAATGCCCTTGTCATATATANNNNNNNNNNNNNNNNNNNNNGA 381  
QY 191 AspAspProGluGluValGluGluYsrGlnYsrGlnYsrGlnYsrGlnYsrGlnYsr 210  
DB 382 GACCATCTCTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 441  
QY 211 GluSerArgProProArgGlyPheProSerAspYsrIlePheGluAlaIleSerSerMet 230  
DB 442 GAAAGCCGCCCACTCGGAATTTCTTGTATAAATTTTGAAGCATTTCTCATATG 501  
QY 231 PheProAspYsrGlyThrAlaGluGluLeuYsrGlyTyrGlyGluLeuThrGluGln 250  
DB 502 TTTCAGATTAAGGACAGACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 561  
QY 251 GlnLeuProGlyAlaLeuProProGluCysThrProAsnIleAspGlyProAsnAlaYs 270  
DB 562 CAGCTCCAGCGGACCTCTCTGAATGTACCCCAACATAGTGAACCAATGCTTAA 621  
QY 271 SerValGlnArgGluGlnSerLeuHisSerPheHisThrLeuPheCysArgArgCysPhe 290  
DB 622 TCTGTTCAAGAGAGCAAGCTTACCTCTTCAATAGCTTTCTGTAGCGATGTTT 681  
QY 291 YsrTyrAspCysPheLeuHisProPheHisAlaThrProAsnThrTyrGlyAspGlyAsn 310  
DB 682 AAATATGAGCTCTCTTACATCTTTTCAATGCTTTTCTGTAGCGATGTTT 741  
QY 311 ThrGluThrAlaLeuAspAsnYsrProCysGlyProGlnCysTyrGlnHisLeuGluGly 330  
DB 742 ACAGAAAAGCTCTTACAGCAAGCTTGTGACACAGTGTTCACAGCATTTTGAAGGA 801  
QY 331 AlaYsrGluPheAlaAlaAlaLeuThrAlaGluYsrGlyIleYsrThrProProYsrArgPro 350  
DB 802 GCAGAGAGTTTGTCTGCTCTCAACCGCTGAGCGGATTAAGAACCCCAACCAAGCTCCA 861





```

143 TGGAAAAGAAAGTGAATCTGATTCATGCCGACTTCGACAACTTAAAGCGTTCAGACA 202
QY 35 A1aapgluVallybSerMetPheSerSerAsnArgLysIleuLglnArgThrglu 54
Db 203 AATATGGGTGCAAGGCTTTGTATGTGCGCAAAATTTTGCAGAAAGGTTCAAGAAAAACCCGAC 262
QY 55 IleuAsnGlnGlnThrgLysGlnArgGllEglnProValHisIleuLeuThrsVal 74
Db 263 ATCTCAATGAAGAAATGGAAGAGCTTCGTGTCACACTGTTCACTCAATGAAGCTGTG 322
QY 75 SerSerLeuArgGlyThzArgLysCysSerValThrsSerAspLeu---AspPheProThr 93
Db 323 AGTGAACACCCCTTTCTCAAAAGTGTACATAGAGAGCATTTCCCGGAGTTTGAAC 382
QY 94 GlnValIlePheLeuLysThrLeuAsnAlaValAlaSerValProIleMetYrsTerTrp 113
Db 383 CAACATATGTATATGAGTCACTGAACACAGTTGCAATGGTCCCATCATGTATTCCTGG 442
QY 114 SerProLeuGlnGlnAsnPheMetValGlnAspGlnThrValLeuHisAsnIleProThr 133
Db 443 TCCCTCTCCACAGAACTTATAGTGAAGATGAGACGGTTTGTGCAATATCTCCAC 502
QY 134 MetGlyAspGlnValLeuAspGlnAspGlyThrPheIleGlnGlnLeuIleLysAsnYr 153
Db 503 ATGGAGATGAGATGAAGAAAGAGATGAGACTTTTATGAGAGCTGATCAATACAT 562
QY 154 AspGlyLysValHisGlyAspArgGlu-----CysGlyPheIleAsnAspGln 169
Db 563 GATGGGAAAGTCAATGCTGAGAAAGATGATCCCTGATCCGTTCTGATGTAGTGTCT 622
QY 170 IlePheValGlnLeuValAsnAlaLeuGlnYrThzAsnAspAsp--- 185
Db 623 GTTTTCTGAGATGTGTGATGCTCCCTGAATCAGTACTCAGATGAGAGAGAGAGAGCGCAC 682
QY 186 -----AspAspAspArgGlyAspAspProGlnGlnArgGlu 197
Db 683 AATGACACTCAGATGAGAAAGAGAGATGACAGCAAGAAATCTGCACTAACAGAAAG 742
QY 198 GlnLysGlnLysAspLeuGlnAspHisValArgAspAspLysGlnSerArgProAspArgLys 217
Db 743 AGAAAGCCGATGCTATGTAAGGCAACAA-----AAGAGTTCCAG-----AAACAG 790
QY 218 PheProSerAspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlnThzAla 237
Db 791 TTCCCAATGACATGATGATCTTCAGTGCATGCTCAATGCTCCCTGAGAGATGTGTCCCA 850
QY 238 GlnGlnLeuLysGlnLysYrThzLysGlnLeuThrgLysGlnLeuProGlyAlaLeuPro 257
Db 851 GATGACATGAGAGAGAGGATGACAGAACTTACAGAGATGTCAGACCCCAATGCACTTCCC 910
QY 258 ProGlnCysThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSer 277
Db 911 CCTCAGTGCACACCAACATGATGCCCCCAATGCAAGTCTGTGACAGCGGAGCAATCT 970
QY 278 LeuHisSerPheHisThrLeuPheCysValArgGlyCysPheLysYrThzAspCysPheLeuHis 297
Db 971 CTGCACTCCTTCCACACACTTTTGTGCGCGCTCTTTAATAGACATGCTTCTTCCAC 1030
QY 298 ProPheHisAlaThzProAsnThzYrThzLysArgLysAsnThzGlnThzAlaLeuAspAsn 317
Db 1031 CCTTTTCATGCCACCCCTATATATATTAACCAAGAAATTAAGAAATCAAGTTGAACA 1090
QY 318 LysProCysGlyProGlnCysYrGlnHisLeuGlnGlnAlaLysGlnPheHisAlaAla 337
Db 1091 GAACCATGTGCAAGACTGCTTCTTTGCTGGAAGGAGCAAGAGATATGCC----- 1144
QY 338 LeuThzAlaGlnArgIleLysThzProProLysArgProGlnGlyAlaArgArgArgLysArg 357
Db 1145 -----ATGCTCCAAACCCCGCTCAAGTGTCTGTGTGCTGCGGAGAAAG 1192
QY 358 LeuProAsnAsnSerSerArgProSerThzProThrIleAsn---ValLeuGlnSerLys 376

```

```

1193 CACACATATGTCAGTCTTCTGTCACATGCTCAGCCTCTGCTGCTGAGACTA 1252
QY 377 AspThzAspSerAspArgGlnAlaGlyThzArgThrgLysGlnAsnAspLysGln 396
Db 1253 GAAGAGACAGTGAAGAGAGACAGAGC----- 1279
QY 397 GlnGlnGlnLysLysAspGlnThzSerSerSerSerGlnAlaAsnSerArgCysGlnThz 416
Db 1280 -----AATGACTGGGCTTCCAGTTCTTCAAGAGCTTACTCTGCTGTAGACT 1327
QY 417 ProIleLysMetLysProAsnIleGlnProPro----- 427
Db 1328 CCCACAAAGAGAGGCTATGTCAGACCCCACTCAACTGTGCGTATGAGAGACCCCTG 1387
QY 428 GlnAsnValGlnThzPheGlyAlaGlnAlaSerMetPheArgValLeuIleGlyThzYr 447
Db 1388 GAGCTGTGAGATGAGCTGGGGCTGAAATCTCTTTTCAAGTCTTCCATGCGACCTAC 1447
QY 448 TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThzThzCysArgGlnValYr 467
Db 1448 TTCAACACTTCTGTTCAATACAGGCTTCTGAGGACAAAGACSTGCAAGAGGCTTT 1507
QY 468 GlnPheArgValLysGlnSerSerIleIleAlaProAlaProAlaGlnAspValAspThz 487
Db 1508 CAGTTTGCAGTCAAGAA-----TCACTTATCTGAAAGCTGCCAAGAGAGCTCATGAA 1564
QY 488 ProProAspGlyLysLysLysArgLysHisValArgLeuThzAlaHisCysArgLysIleGln 507
Db 1565 CCTTCACAGAAAGAAAGAAAGAGACAGATGTGGGGCTGCACTGCGAGAGAAATTCAG 1624
QY 508 LeuLysLysAspGlySerSerAsnHisValTyrAsnYrGlnProCysAspHisAspArg 527
Db 1625 CTGAAGAAAGTACTTCTTCCACAGAGTGAACATACCAACCTTGGAGACCAACCAAGC 1684
QY 528 GlnProCysAspSerSerCysProCysValIleAlaGlnAsnPheCysGlnLysAspCys 547
Db 1685 CGCCCTGTGACAGCACTGCCCCCTGATCATGACTCAAGAAATTTCTGAGAAAGTTCTGC 1744
QY 548 GlnCysSerSerGlnCysGlnAsnArgPheProGlyCysValGlyLysAlaGlnCysAsn 567
Db 1745 CAGTGCACACCAAGACTGTGCAAGTGTGTTCCCTGCTGCTGCTGCTGTAAGACCCAGTGCAT 1804
QY 568 ThrLysGlnCysProCysYrThzLeuAlaValArgGlnCysAspProAspLeuCysLeuThz 587
Db 1805 ACCAAGCAATGCTCTGCTATGTGCAAGTGAAGATGACCTGACCTGTGTCTACCC 1864
QY 588 CysGlyAlaAlaAspHisIleThzAspSerLysAsnValSerCysLysAsnCysSerIleGln 607
Db 1865 TGTGGGGCTCAGAGCACTGGGAGCTGCAAGGTGTCTCTGTAAATACTGACAGATCCAG 1924
QY 608 ArgGlySerLysLysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIlePhe 627
Db 1925 CGTGACTTAAGAAAGCACTGCTGCGCCCCCTGTGATGTGCGCGAGTGGGGCACTTTC 1984
QY 628 IleLysAspProValGlnLysAsnGlnPheIleSerGlnYrCysGlyGlnIleIleSer 647
Db 1985 ATAAAGAGCTGTGCAAGAAAGCAATTCATTTCTGAATACGTGGGAGGTCAATCTCT 2044
QY 648 GlnAspGlnAlaAspArgArgGlyLysValTyrAspLysYrMetCysSerPheLeuPhe 667
Db 2045 CAGAGATGAGGCTGATGCGAGGAAAGGTCTATGACAAATACATGTCCAGGCTTCTTTC 2104
QY 668 AsnLeuAsnAsnAspPheValValAspAlaThzArgLysGlyAsnLysIleArgPheAla 687
Db 2105 AACCTCATTAATGATTTTGTGTGATCTCTCTCGAAAGAGAAACAAATTCGATTTTCA 2164
QY 688 AsnHisSerValAsnProAsnCysYrAlaLysValMetMetValAsnGlyAspHisArg 707
Db 2165 AATCATTTAGGAATCCCACTGTATGCCAAAGGTGATGTGAAATGAGAACCATGG 2224
QY 708 IleGlyIlePheAlaLysArgAlaIleGlnThzGlnGlnLysPhePheAspYrArg 727
Db 2225 ATTGGGATCTTTGCCAAGAGGCAATTCACACTGGCGAAGGCTCTTCTTGAATTACAGG 2284

```

Qy 728 TySerGlnAlaAaPalaLeuTyTyValGlyIleGluArgGluMetGluIle 745  
Db 2285 TACGCCAAGCTGATGCTCTCAAGTACGTGGGATCGAAGAGGACCGACGCTC 2338  
RESULT 9  
US-11-266-748A-26094  
Sequence 26094, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Hartin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcription Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 26094  
LENGTH: 4640  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-26094  
Alignment Scores:  
Pred. NO.: 3,14e-210 Length: 4640  
Score: 2603.50 Matches: 490  
Percent Similarity: 76.84 Conservative: 92  
Best Local Similarity: 64.64 Mismatches: 123  
Query Match: 64.38 Indels: 53  
DB: 8 Gaps: 10  
US-10-773-302-2 (1-746) x US-11-266-748A-26094 (1-4640)  
Qy 15 TTPAAGYSAAGVALYVSEGLNYTMTARGLAAGGGINLEUYAARGPHEARG 34  
Db 165 TGGAAAAAATAAATCTGAATATGATGCACTTTCACAACTTAAACGGCTTCAAGCA 224  
Qy 35 ALaAaPalaLeuTySerGlnAlaAaPalaLeuTyTyValGlyIleGluArgGlu 54  
Db 225 AATATGGGTGCAAGGCTTGTATGTGGCAATTTTGCAGAGCTTCAAGAAAAAACCAG 284  
Qy 55 ILeuAaGlnGluTyrIleGlnArgGlyIleGlnProValHisIleuGluArgGlu 74  
Db 285 ATCTCATGAAAGATGAAAGAAAGCTGTGTCACCACTGTTCACTCAATGAAGCTGTG 344  
Qy 75 SerSerLeuAaGlyTyrArgGluGlySerValThrSerAaPalaLeu---AaPheProThr 93  
Db 345 AGTGACACCCCTTTCTCAAAAAGGTATCCATATGAGAGCATTTTCCGGGATTTGCAAGC 404  
Qy 94 GlnValIleProLeuTySerGlnAlaAaPalaLeuTyTyValGlyIleGluArgGlu 113  
Db 405 CAACATATGTTAATAGGCTGACTGAACACAGTTGATGGTTCCCATCATGATATTCCTGG 464  
Qy 114 SerProLeuGlnGluAaPheMetValGluAaPalaLeuHisAaPalaLeuProTy 133

Db 465 TCCCTCTCCAAAGCACTTATATGTTGAAGATGACGGTGTTCGCAATATTCCTAC 524  
Qy 134 MetGlyAaPalaLeuAaPalaLeuAaPalaLeuTyTyValGlyIleGluArgGlu 153  
Db 525 ATGGAGATGATAGTGAAGAAAGAAAGATGACATTTTATGAGAGCTGATCATATCTAT 584  
Qy 154 AaPalaLeuTyTyValGlyIleGluArgGlu-----CysGlyPheIleAaPalaLeu 169  
Db 585 GATGGAAAGTCCATGCTGAAAGAAAGATGATCCCTGATCCGTTCTGATATGATGATGCT 644  
Qy 170 ILeuValGluLeuValAaPalaLeuGlyIleGluTySerAaPalaLeu----- 185  
Db 645 GTTTTCTGAGTTGTGATGATGCTTAAATCATGATCTCAATGAGATGAGTATGATGATGCT 704  
Qy 186 -----AaPalaLeuAaPalaLeuAaPalaLeuAaPalaLeuAaPalaLeuAaPalaLeu 197  
Db 705 AATGACACCTCAATGAAAGCAAGATGACAGCAAAAGATTCGCAAGTAAAGAAAG 764  
Qy 198 GlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 217  
Db 765 AGAAAGCAATGCTATTTGAAGGACAA-----AAGAGTTCAG-----AAACAG 812  
Qy 218 PheProSerAaPalaLeuPheGlnAlaIleSerSerPheProAaPalaLeuTyTyVal 237  
Db 813 TTCCCAATGACATGATCTTCAATGATGCTCAATGCTCAATGCTCAATGCTCAATGCTCA 872  
Qy 238 GlnGluLeuTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySer 257  
Db 873 GATACATGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 932  
Qy 258 ProGlnTySerProAaPalaLeuAaPalaLeuAaPalaLeuAaPalaLeuAaPalaLeu 277  
Db 933 CCTGAGTGCACACCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 992  
Qy 278 LeuTySerPheIleGlnAlaIleSerSerPheProAaPalaLeuTyTyValGlnTySer 297  
Db 993 CTGCACTCTTCCACACATCTTTCCTGCGGCTGCTTAAATGATGATGATGATGATGATGAT 1062  
Qy 298 ProPheIleAaPalaLeuTyTyValGlnTySerGlnTySerGlnTySerGlnTySerGln 317  
Db 1053 CTTTTCATGCAACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112  
Qy 318 TySerProGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 337  
Db 1113 GAACCATGTCGACACATGCTGCTTTCCTGTCGAAAGAGCAAGAGTATGATGATGATGATGAT 1166  
Qy 338 LeuThrIleGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 357  
Db 1167 -----ATGCTCCACACCCCGCTCCAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1214  
Qy 358 LeuProAaPalaLeuTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 376  
Db 1215 CACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1274  
Qy 377 AaPalaLeuTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySer 396  
Db 1275 GAAGAGACAGTGCACAGGACACAGG----- 1301  
Qy 397 GlnGluGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySer 416  
Db 1302 -----AATGACTGGGCTCCAGTTCCTCAAGGCTAATCTGCTGCTGCTGCTGCTGCT 1349  
Qy 417 ProIleTySerTySerProAaPalaLeuProPro----- 427  
Db 1350 CCCACAAACAGAGGCTATGCTGACGCCACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1409  
Qy 428 GlnAaPalaLeuTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 447  
Db 1410 GAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1469  
Qy 448 TySerAaPalaLeuTySerGlnTySerGlnTySerGlnTySerGlnTySerGlnTySerGln 467

Db 1470 TTCAACAACTCTGTTCAATAGCAGGCTTCTGGGACCAAGACGTCAGACGCTTT 1529  
Qy 466 GtupheargValylgluserSerIlelleaIProAlaGluAspValleuphr 487  
Db 1530 CAGTTTGAGTCAAGAA--TCACTTATCTGAAGCTGCAACAGATAGCTGTAAC 1586  
Qy 488 ProProArgLysLysLysArgLysHisArgLeuThrAlaHisLysArgLysIleGln 507  
Db 1587 CCTTCACAGAAAGAAAGAAAGAAAGCAAGATTTGGGCTGCACATCGAGAAAGATTCCG 1646  
Qy 508 LeuLysLysAspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArg 527  
Db 1647 CTGAAGAAAGATTAATCTTCCACACAGAGTGTAACTAACCAACCTGCGACACCAAGAC 1706  
Qy 528 GlnProCysAspSerSerCysProCysValIleleAlaGlnAspHisCysLysPheCys 547  
Db 1707 CGCCCTGTGACAGCACTGCCCCCTGCAGCATGACTCAGAAATTTCTGTAGAGATTCGC 1766  
Qy 548 GlnCysSerSerGluCysGlnAsnArgPheProGlyCysArgCysLysValIleGlnCysAsn 567  
Db 1767 CAGTCACACCCAGACTGTGAGATGCTTTCCCTGGCTGTGCTGTAAGCCAGTCAGAT 1826  
Qy 568 ThrLysGlnCysProCysTyrLeuAlaValArgGluCysAspProAspLeuCysLeuThr 587  
Db 1827 ACCAAGCAATGTCTTGTATCTGGCAGTGCAGAGATGACCTGACCTGTGTCTCAC 1886  
Qy 588 CysGlyAlaAlaAspHisTTrpAspSerLysAsnValSerCysLysAsnCysSerIleGln 607  
Db 1887 TGTGGGGCTCAGACACTGGGACTGCAAGGTGGTTCCTGTGTAAGAACTGCGACATCCG 1946  
Qy 608 ArgGlySerLysLysHisLeuLeuLeuAlaProSerAspValAlaGlyTrrGlyLysPhe 627  
Db 1947 COTGACCTTAAGAGCACTGCTGCTGGCCCCCTCTGATGTGCGGATGGGACCTTC 2006  
Qy 628 IleLysAspProValGlnLysAsnGluPheIleSerGluTyrCysGlyGluIleLysSer 647  
Db 2007 ATTAAGAGACTGTCTGCAAGAAAGCAATTCATTTCTGAATATCTGTGTGAGCTCATCTT 2066  
Qy 648 GlnAspGluAlaAspArgArgGlyLysValTyrAspLysTyrMetCysSerPheLeuPhe 667  
Db 2067 CAGAGTGAAGCTGATCGACGCGGAAAGCTTATGCAATATCATCTCCAGCTTCTCTTC 2126  
Qy 668 AsnLeuAsnAsnAspPheValValAspAlaThrArgLysGlyAsnLysIleArgPheAla 687  
Db 2127 AACTCATAATATGATTGTGTAGTGAATCTCTCGAAAGAAACAAATTCGATTGGA 2186  
Qy 688 AsnHisSerValAsnProAsnCysTyrAlaLysValMetMetValAsnGlyAspHisArg 707  
Db 2187 AATCATTGAGTGAATCCCAACTGTTATGCAAAAGTGTGATGTGATGAGACATCGG 2246  
Qy 708 IleGlyIlePheAlaLysArgAlaIleGlnThrGlyGluLysPhePheAspTyrArg 727  
Db 2247 ATTGGGATCTTTGCCAAGGGGCAATTCAGCTGCGAAGAGCTCTTCTTGATTACAGG 2306  
Qy 728 TyrSerGlnAlaAspAlaLeuLysTyrValGlyIleGluArgGluMetGluIle 745  
Db 2307 TACAGCCAGCTGATGCTCTCAAGTACGTGGGATCGAAGGAGACCGACGCTC 2360

## RESULT 10

US-11-266-748A-186457/c  
; Sequence 186457, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIOR FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03

; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: Patent version 3.3  
; SEQ ID NO 186457  
; LENGTH: 2074  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-11-266-748A-186457  
  
Alignment Scores:  
Pred. No.: 4,07e-144 Length: 2074  
Score: 1817.50 Matches: 369  
Percent Similarity: 55.3% Conservative: 1  
Best Local Similarity: 55.2% Mismatches: 7  
Query Match: 44.9% Indels: 293  
DB: 8 Gaps: 2  
  
US-10-773-302-2 (1-746) x US-11-266-748A-186457 (1-2074)  
Qy 186 AspAspAspAspAspGlyAspAspProGluGluArgGluLysGluLysAspLeuGluAsp 205  
Db 2069 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2010  
Qy 206 HisArgAspAspLysGluSerArgProProArgLysPheProSerAspLysIlePheGlu 225  
Db 2009 CACCGAGTATTAAGAAAGCCGCCACCTCGAATATTCCTTCTGATTAATATTTTGA 1950  
Qy 226 AlaIleSerSerMetPheProAspLysGlyThrAlaGluLysGluLysGlyLysTyrLys 245  
Db 1949 GCCATTTCTCAATGTTTCCAGATTAAGGACACAGAGAAAGAACTTAAGAAATATAA 1890  
Qy 246 GlnLeuThrGluGlnLysPheProGlyAlaLeuProProGluCysThrProAsnIleAsp 265  
Db 1889 GAATCTCACCGAAGCAGAGCTCCAGGCGCACTTCTCTGATATATACCCCAACATAGAT 1830  
Qy 266 GlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSerPheHisThrLeuPhe 285  
Db 1829 GAGCCAAATGCTAATCTGTTCAAGAGAGGAAAGCTTACCTCTTTCATACGCTTTTC 1770  
Qy 286 CysArgArgCysPheLysTyrAspCysPheLeuHisProPheHisAlaThrProAsnThr 305  
Db 1769 TGTAGGCAATTTTAATATAGTCTGCTTCTTCAATCTTTTCATGCAACCCACACT 1710  
Qy 306 TyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCysGlyProGlnCysTyr 325  
Db 1709 TATTAAGCGAAGAAACAGAAACAGCTCTAGACAAACCTTGTGACCAAGTGTAC 1650  
Qy 326 GlnHisLeuGluGluAlaLysGluPheAlaAlaLeuThrAlaGluArgIleLysThr 345  
Db 1649 CAGCATTTGAGGAGGAGCAAGAGATTTGCTGCTCTCTCAACGCTGAGGAGTAAAGCC 1590  
Qy 346 ProProLysArgProGlyArgArgArgGlyLysLeuProAsnAsnSerSerArgPro 365  
Db 1589 CCACCAAAAGCTCCAGAGAGCGCGCAAGAGAGAGAGGCTTCCCATTAACAGTACAGGCCC 1530  
Qy 366 SerThrProThrIleAsnValLeuGluSerLysAspThrAspSerAspArgGluAlaGly 385  
Db 1529 AGCACCCCAACATTAATGTCTGTAAGATCAAGAGATCAAGACGTGTATAGGAGAGAGG 1470  
Qy 386 ThrGluThrGlyGlyGluAsnAsnAspLysGluGluGluGlyLysArgGluThrSer 405





206 HsArGAspAspLysGluSerArgProProArgLysPheProSerAspLysIlePheGlu 225  
66 CACCGAGATGATTAAGAAAGCCGCCACTCGGAATTTCTTCTGATTAATTTTGA 125  
226 AAlIeSerSerMetPheProAspLysGluThrAlaGluGluLeuLysGluLysTyrls 245  
126 GCATTCTCTCAATGTTTCCAGTAAAGGCGACAGGAGAACTAAAGAAATATATA 185  
246 GluLeuThrGluGluGluLeuProGluAlaLeuProProGluLysThrProAsnIleAsp 265  
186 GAACCTACCGAAGAGAGCTCCAGCGGACTTCTCTGAAATGTAACCCCAACATAGAT 245  
266 GlProAsnAlaLysSerValGluArgGluGlnSerLeuHisSerPheHisThrLeuPhe 285  
246 GCACCAATGCTAAATCTTTCTAGAGAGGCAAGCTTCACTCTTCTATAGCTTTTC 305  
286 CybArGArgCybPheLysTyrlsPheLysHisProPheHisAlaThrProAsnThr 305  
306 TGTAGGCGATGTTTAAATATGACTGCTTCTTAACATCTTTTCATGCAACCCAACT 365  
306 TyrlsArgLysAsnThrGluThrAlaLeuAspAsnLysProCysGluProGluCysTyrls 325  
366 TATAAGCGGAGAACACAGAAACAGCTTACACAAACCTTGTGACACAGAGTGTAC 425  
326 GlnHisLeuGluGluValAlaLysGluPheAlaAlaAlaLeuThrAlaGluLysThr 345  
426 CAGCATTTGAGAGGAGCAAGAGAGTGTGCTGCTCTCAACGCTGAGGAGATTAAGACC 485  
346 ProProLysArgProGluLysArgArgGluLysArgLeuProAsnAsnSerSerArgPro 365  
486 CCAACCAAAAGCTCCAGAGAGCGCGAGAGAGAGCGCTTCCCAATTAACGTAGAGCGCC 545  
366 SerThrProThrIleAsnValLeuGluSerLysAspThrAspSerAspArgLysAlaGly 385  
546 AACACCCCAACCATTAATGCTGCTGAGATCAAGAGATACAGACAGTGTAGGAGAGAGG 605  
386 ThrGluThrGluLysGluAsnAsnAspLysGluGluGluLysLysAspGluThrSer 405  
606 ACTGAAACCGGGGAGAGAGACATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665  
406 SerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMetLysProAsnIleGlu 425  
666 ACTCCTCTGAAAGCAATTTCTGCTGCTCAACCAACCAATTAAGTGAAGCAATATGAA 725  
426 ProProGluAsnValGluThrSerGluAlaGluAlaSerMetPheArgValLeuIleGly 445  
726 CCTCCTGAGAAATGTGAGAGTGTGCTGAAAGCTCAATGTATTAGATCTCTCAATTGCG 785  
446 ThrTyrlsAspAsnPheCysAlaIleAlaArgLeuIleGlyThrLysThrCysArgGln 465  
786 ACTTACTATGACAATTTCTGTGCTGCTAGGTTAATTGGAGCCAAACATGTAGACG 845  
466 ValTyrlsGluPheArgValLysGluSerSerIleIleAlaProAlaProAlaGluAspVal 485  
846 GTGTATGAGTTTATGAGTCAAGATATTAACATCAATAGCTCCAGCTCCCGCTAGAGATG 905  
486 AspThrProProArgLysLysLysArgLysHis----- 496  
906 GATACTCTCTCAAGGAAAAAGAGAAACACCGCTTTAGAGAGATGCTTCAAAATT 965  
496 ----- 496  
966 TGGTGAATTATTTCTGACCAAGTATGCTTGTGGTTATATATGCTGAAATTAAGA 1025  
496 ----- 496  
1026 GGAAGTATCAGTGTATCATCATCATGATTAATCAAGGAGACACACTGTGCTTT 1085  
496 ----- 496  
1086 GTGACAGATTGGAGCTCAGAGTTCTGACAGAGATCATGATGAACACAGAAATATAC 1145  
496 ----- 496

1146 GTCTTGCCACACTGGGGGAGCGTGAAGCCATCAAGCTGTGTTAGTGTCTTAACATGATG 1205  
496 ----- 496  
1206 GCAAAAAATAAACTGGGAGGCGATGACTTTGGCTTTAAGCATTCCTGTATTCATGCA 1265  
496 ----- 496  
1266 CAGATGATCCAGTTTAAGCAGCTCTGTGTGATTTGTAGCTTCCCGAGAAATTTGGTT 1325  
497 ----- 497  
1326 TAAATTTCTTGTGTTTGTGAGAGGTGTGGCTGACACTGAGAAAGATACAGCTGAAA 1385  
510 Lys----- 510  
1386 AAGGCTTAGCATCTTTCATTCCTCTCATTTATTAAGCTTAACAAATTCGTGTTTGTCA 1445  
510 ----- 510  
1446 TTTGTTTAAATAGTAACCTGAGATTAAGCTTGGTTATTTCTAATATATATGCTTGT 1505  
510 ----- 510  
1506 ATGTTAAACAAATTGTAATGAGAGAGCTCTACTAGTGAATCCCAAGTCAATTCCTGTG 1565  
510 ----- 510  
1566 TAAAGTAGAGCTGACTGGAAGAGTGTAGTCCAGTCACTTGAAGAGAGAGTCTTTGGG 1625  
510 ----- 510  
1626 ATGGACTGAGATGTAAGCGGTGCTTGTATGATGAAGTCAATGAACCTTGGGGGTGGAGAG 1685  
510 ----- 510  
1686 TATTTTTCATCGTTGTGGAACCGGAATATGTATATGTTATGATTCAGTAAAGGC 1745  
510 ----- 510  
1746 AGGCAAGTACACTCCACAGATGATGAGAAAGATGACTGCTGTTAATGTATGATGATC 1805  
511 ----- 511  
1806 GTTCCATCTCCCTGATTCATTTGCTGATGATGATTTATCTTAATCTGAGAGCGCTCTC 1865  
514 TAsnHisValTyrlsAsnTyrlsProCysAspHisAspArgGlnProCysAspSerSerCys 534  
1866 TAAACATGTTTACACATATCAACCTGTGATCATCAAGGAGCTTGTGACAGTTGCTG 1925  
534 sProCysValIleAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlu 554  
1926 CCTTGTGTGATGAGACA- AATTTTGTGAAAGATTTGTCAATGTAGTTGAGAGTGA 1984  
554 nAsnArgPheProGluCysArgCys 562  
1985 GTATTTGTGCTTGTGATGCAATTTGC 2009  
RESULT 12  
US-11-266-748A-98588/c  
; Sequence 98588, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcription Microarray Technology and  
; FILE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; CURRENT FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2

```

; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 48396
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 98588
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (86)..(86)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (169)..(169)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (362)..(362)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (367)..(367)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY:
; LOCATION: (468)..(468)
; OTHER INFORMATION: n is a, c, g, or t
; OTHER INFORMATION: n is a, c, g, or t
US-11-266-748A-98588

Alignment Scores:
Pred. No.: 1.01e-141 Length: 1140
Score: 1785.00 Matches: 350
Percent Similarity: 92.54 Conservative: 6
Best Local Similarity: 90.98 Mismatches: 21
Query Match: 44.1% Indels: 10
DB: Gaps: 0

US-10-773-302-2 (1-746) x US-11-266-748A-98588 (1-1140)
QY 274 ArgGluGlnSerLeuHisSerPheHisThrLeuPheCysValGlyArgCysPheLeuYTrAsp 293
DB 1138 AGAGGCAAGAGTAACTCTTCATGCAACCACTTTCTGTAGCGGATGTTTAAATATGAC 1079
QY 294 CysPheLeuHisProPheHisValThrProAsnThrYTrLysArgLysAsnThrGluThr 313
DB 1078 TGCTTCTTACATCTTTTCATGCAACCACTTTCTGTAGCGGATGTTTAAATATGAC 1079
QY 314 AlaLeuAspAsnLysProCysGlyProGlnCysYTrGlnHisValGluGluValAlaLysGlu 333
DB 1018 GCTTAGACCAACCAACTTGTGACCAACAGTGTACAGCATTTTGAAGGAGCAAAAGAG 959
QY 334 PheAlaAlaAlaLeuThrAlaGluArgLysLeuThrProProLysArgProGlyGlyArg 353
DB 958 TTGTGCTGTGCTCTCAACCGCTGAGGAGTAAGAACCAACCAACGTCACGAGGCGCG 899
QY 354 ArgArgGlyArgLeuProAsnHisSerSerArgProSerThrProThrIleAsnValLeu 373
DB 898 AGAAGAGACCGCTTCCCAATACAGTAGACAGGCCCAACCCCACTTATATGTGCTG 839
```

```

QY 374 GluSerLysAspThrAspSerAspArgGluAlaGlyThrGlnThrGlyGlyGluAsnAsn 393
DB 838 GAATCAAAAGATACAGCACTGATATGGAAGCAGGAGCTAAACGGGGGAGAGAACAT 779
QY 394 AspLysGluGluGluGlyLysAspGlyThrSerSerSerSerGluAlaAsnSerArg 413
DB 778 GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
QY 414 CysGlnThrProIleLysMetLysProAsnIleGluProGluAsnValGluTrpSer 433
DB 718 TGTCAAACCAACCAATTAAGATGAAGCCAAATTTGAACTCTCGAGAAATGCGAGT 659
QY 434 GlyAlaGluLysSerMetPheArgValLeuIleGlyThrYTrLysAspAsnPheCysAla 453
DB 658 GGTGCTAAGCTCAAGTTTAAAGTCTCAATGCGCACTTACATGACAAATTTCTGTGCC 599
QY 454 IleAlaArgLeuIleGlyThrLysThrCysArgGlnValYTrGluPheArgValLysGlu 473
DB 598 ATTGCTAGCTTAAATTGGAGACCAAAACATGTAGACAGGTGTATGAGTTTAGAGTAAAGAA 539
QY 474 SerSerIleIleAlaProAlaProAlaGluAspValAspThrProArgLysLysLys 493
DB 538 TCTAGCATCATAGCTCCAGCTCCGCTGAGGATGTGATCTCTCCAGAGAAAGAAAG 479
QY 494 ArgLysHis-ArgLeuThrAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySe 513
DB 478 AGGAACAACCGGTGTGGGCTGCACCTGACAGAAATACAGCTGAGAAAGACGGCTC 419
QY 513 rSerAsnHisValYTrAsnYTrGlnProCysAspHisProArgGlnPro-CysAsp-Ser 532
DB 418 CTCTAACCATGTTTACACATCTTCAACCTGTGATCATCCAGGAGCTTGNTGACAGT 359
QY 533 SerCysPro-CysValIleAlaGlnAsnPheCys-GluLysPheCysGlnCysSerSer 551
DB 358 TCGTGCCCTTGTGTGATGACCAAAATTTTGTGTAAGATTTTGTCAATGATGATCA 299
QY 552 GluCysGlnAsnArgPheProGlyCysArgCysLysValGlnCysAsnThrLysGlnCys 571
DB 298 GAGGTCAAAACCGCTTCCGGATGCGCTGCACACACAGTGCACACACACACAGTGC 239
QY 572 ProCysYTrLeuAlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAla 591
DB 238 CCGTCTACCTCGCTGCCAGAGTGTGACCTTCACTCTCGTCTACTGTGAGGGGCT 179
QY 592 AspHisThrAspSerLysValAsnValSerCysLysAsnCysSerIleGlnArgLysSerLys 611
DB 178 GACCATTTGACCACTTAATAATGTCTCTGCAAGAACTGCAAGTATTCAGCGGGCTCCA- 120
QY 612 LysHisLeuLeuLeuAlaProSerAspValAlaGlyTrpGlyIle-PheIleLysAspPr 631
DB 119 AAGCATCTATTGCTGCACCATCTGACTGGGAGAGCTTGGGGGATTTTATTCAAGATCC 60
QY 631 oValGlnLysAsnGluPheIleSerGlnYTrCysGlyGluIleIle-SerGlnAspGlu 650
DB 59 TGGGCAAGAAATGAGAT--CTCTAAATATCTGGGGGAAAGATTATTTTCCCAATGA 3

RESULT 13
US-11-266-748A-151399
; Sequence 151399, Application US/11266748A
; Publication No. US2006013463A1
; GENERAL INFORMATION:
; APPLICANT: Hartlin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcription Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 5815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; PRIOR FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
```

PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 151399  
LENGTH: 1140  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (673)..(673)  
FEATURE:  
NAME/KEY: n is a, c, g, or t  
LOCATION: (774)..(774)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (779)..(779)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: (972)..(972)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1055)..(1055)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-151399

Alignment Scores:  
Pred. No.: 1,01e-141 Length: 1140  
Score: 1785.00 Matches: 350  
Percent Similarity: 92.5% Conservative: 6  
Best Local Similarity: 90.9% Mismatches: 21  
Query Match: 44.1% Indels: 10  
DB: 8 Gaps: 0

US-10-773-302-2 (1-746) x US-11-266-748A-151399 (1-1140)

QY 274 ArgGluGlnSerLeuHisSerPheHisThrLeuPheCysArgArgCysPheLeuSerThrArgP 293  
DB 3 AGAGAGCAAGCTTACCTCTTTCATAGCGCTTTCTGAGCGCATGTTTAATATGAC 62  
QY 294 CysPheLeuHisProPheHisAlaThrProAsnThrTyrLeuArgGlyAsnThrGlnThr 313  
DB 63 TCGTCTTACATCTTTTCATGCAACCCCACTATTAACGGAAGAACACAGAAACA 122  
QY 314 AlaLeuAspAsnLeuProCysGlyProGlnCysTyrGlnHisLeuGlnGlyAlaLeuGln 333  
DB 123 GCTTACGACACAAACCTTGTGACACAGTGTTCACGCACTTTGAGGAGACAAAGAG 182  
QY 334 PheAlaAlaAlaLeuThrAlaGlnArgGlyLeuThrProProCysArgProGlyGlyArg 353  
DB 183 TTGCTGCTGCTCTCAACCGCTGAGCGGATAAAGACCCCAAAACGTCACAGAGCGCGC 242  
QY 354 ArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu 373  
DB 243 AGAAGAGGAGCGCTTCCCAATACAGTAGACAGCCCAAGCAACCCCACTTAATATGCTG 302  
QY 374 GluSerLeuAspThrAspSerAspArgGlnAlaGlyThrGlnThrGlyGlyGluAsn 393  
DB 303 GAATCAAGATACAGACAGTAGTAGGAGGAGCGGAGTGAACCGGGGAGAGAACAT 362

QY 394 AspLeuGluGlnGluLeuLeuAspGlnThrSerSerSerSerGluAlaAsnSerArg 413  
DB 363 GATTAAGAGAGAGAGAGAGAGATTAACCTTCAGCTCTCTGAAGCAATCTCGG 422  
QY 414 CysGlnThrProIleLeuMetLeuProAsnIleGluProGluAsnValGluThrSer 433  
DB 423 TGTCAACACCAATTAAGATAGAGCAATATTGAACCTCTGAAATAGTAGAGAGT 482  
QY 434 GlyAlaGluAlaSerMetPheArgValLeuIleGlyThrTyrThrAspAsnPheCysAla 453  
DB 483 GGTGCTGAAGCTCAATGCTTAAAGATCCCTGACCTTACCTAAGACAAATTCGTGTC 542  
QY 454 IleAlaArgLeuIleGlyThrLeuThrCysArgGlnValTyrGluPheArgValLeuGln 473  
DB 543 ATGCTAGATTAAATGGAGCAAAACATGTAGACAGGTGATAGATTAGACTTAAGAGA 602  
QY 474 SerSerIleIleAlaProAlaProAlaGlnAspValAspThrProProArgLeuLeu 493  
DB 603 TCTAGCATCATATGCTCCAGCTCCGCTGAGATGTGATACCTCCCAAGGAAAAAGAG 662  
QY 494 ArgLeuHis-ArgLeuThrAlaAlaHisCysArgLeuIleGlnLeuLeuLeuAspGlySe 513  
DB 663 AGAAGACACACGCTGTGGCTGCACACTGACAGAAAGATACAGCTGAAAAAGACGGCTC 722  
QY 513 rSerAsnHisValTyrAsnTyrGlnProCysAsnHisProArgGlnPro-CysAsp-Ser 532  
DB 723 CTCAACCATGTTTAACTAATCAACCTGTGATATATCACGCGCTTGATGACAGT 782  
QY 533 SerCysPro-CysValIleAlaGlnAsnPheCys--GluLeuPheCysGlnCysSerSer 551  
DB 783 TCGTCCCTTGTGTGATAGACAAATAATTTTGGTGAATAATTTGTCAATGATGATCA 842  
QY 552 GluCysGlnAsnArgPheProGlyCysArgCysLeuAlaGlnCysAsnThrLeuGlnCys 571  
DB 843 GAGTGTCAAAACCGCTTCCGGGATGCCGCTGCAAGACAGTGCACCAACCAAGACAGTGC 902  
QY 572 ProCysTyrIleAlaValArgGluCysAspProAsnLeuCysLeuThrCysGlyAlaAla 591  
DB 903 CGGTGCTACCTGGCTGTCCGAGGTGACCTTGACCTCTGTACTTGTGAGCGCT 962  
QY 592 AspHisThrAspSerLeuAsnValSerCysLeuAsnCysSerIleGlnArgGlySerLeu 611  
DB 963 GACCATTTGACACGTAATAATGTGTCTCGAAGAACTGCAATTCACGGGGCTCCAA- 1021  
QY 612 LysHisLeuLeuLeuAlaProSerAspValAlaGlyThrGlyLeu-PheIleLeuAspPr 631  
DB 1022 AAGCATTAATGCTGACACATCTGACTGGAAAGCTTGGGGGATTTTATTCAGATCC 1081  
QY 631 CysAlaGlnLeuAsnGluPheIleSerGluTyrCysGlyGluIleLeu-SerGlnAspGlu 650  
DB 1082 TGGGCAAGAAATGAGAT--CTCTCAAAATCTGGGGGAAAGATTTATTTCTCAAGTGA 1138

RESULT 14  
US-11-266-748A-175796  
Sequence 175796, Application US/11266748A  
Publication No. US2006013463A1  
GENERAL INFORMATION:  
APPLICANT: Hartin, Paul  
APPLICANT: Johnson, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcription Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0

```
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105485.9
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105484.2
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: US 60/662,276
/ PRIOR FILING DATE: 2005-03-14
/ PRIOR APPLICATION NUMBER: US 60/700,293
/ PRIOR FILING DATE: 2005-07-18
/ NUMBER OF SEQ ID NOS: 483996
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 175796
/ LENGTH: 1011
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (18)..(18)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (64)..(87)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (141)..(141)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (151)..(151)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (212)..(212)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (222)..(222)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (238)..(238)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (251)..(251)
/ OTHER INFORMATION: n is a, c, g, or t
/ OTHER INFORMATION: n is a, c, g, or t
US-11-266-748A-175796

Alignment Scores:
Pred. No.: 1.07e-120 Length: 1011
Score: 1536.00 Matches: 315
Percent Similarity: 92.74 Conservative: 3
Best Local Similarity: 91.88 Mismatches: 14
Query Match: 37.98 Indels: 11
Gaps: 0

US-10-773-302-2 (1-746) x US-11-266-748A-175796 (1-1011)

QY 162 GUCYGGYpHeiLeaSpaSpGluLeuValaAaAlaLeuGlyGln-TY 181
DB 2 GAATCTGGGTTTAAATGATGAATTTTGTGGAGTTGCTAAAGCCCTGGTCAATA 61
QY 181 rAsnaSpaSpaSpaSpaSpaSpaGlyAspaSpProGluGluArgGluGluGlnTy 201
DB 62 TAAANNNNNNNNNNNNNNNNNNNNNNNGAGACATCTCTGAAGAAAGAAAGAAAGCA 121
QY 201 aAspleuGluAspHisArg-AspaSpLyS-gluSerArgProPro-ArgLySpHePro-S 220
DB 122 AGATCTGGAGGATGACCGGANGATGATTAANGAAAGCCGCCCACTTCGGAATTTCTCT 181
QY 220 eRaSpLySIllePhe-GluAlaIleSerSer-MethPro-AspLySgLyThrAla-Glu 238
```

```
DB 182 CTGATTAATTTTGTGAAGCATTTCTCANATGTTCCANGATTAAGGACAGCANGAA 241
QY 239 GluLeuLyS-GluLySgLyLeuThrGluLeuThrGluGlnGluLeuProGlyAlaLeuProP 258
DB 242 GAATTAAGGAAATTAATTAAGATCTCACGGAAGACAGCTCCAGGCGCACTTCTCC 301
QY 258 OGluCyEThrProAsnIleAspGlyProAsnAlaLySgValGlnArgGluGlnSerLe 278
DB 302 TGAATGTACCCCAACCAATGATGACCAATGCTTAATGTTCTCAGAGAGCAAGACCTT 361
QY 278 uHisSerPheHisThrLeuPheCysArgArgCysPheLySgTyAspCysPheLeuHisPr 298
DB 362 ACATCTCTTCATACGCTTTCTGTAGGCGATGTTTAATATGACGCTTCTCACTACC 421
QY 298 oPheHisAlaThrProAsnThrTyTyArgLyAsnThrGluThrAlaLeuAspAsnTy 318
DB 422 TTTTCATGCAACACCCAACTTATTAAGCGAAGAAACAGAAACAGCTCTAGCAACAA 481
QY 318 sProCyGlyProGlnCyTyTyGlnHisLeuGluGlyAlaLySgLeuPheAlaIle 338
DB 482 ACCTTGGAACCAAGGTTACCAAGATTGGAGGAGCAAGAGAGTTGCTGCTCT 541
QY 338 uThrAlaGluArgIleLeuThrProProLySgProGlyGlyArgArgGlyArgLe 358
DB 542 CACCGCTGAGCGATTAAGACCCCAACCAACCTCCAGAGCGCCGAGAAAGAGCGCT 601
QY 358 uProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLyAspTh 378
DB 602 TCCCAATTAACAGTACGAGGCCAGCACCCCACTTAATGTCTGTAATCAAGATAC 661
QY 378 rAspSerAspArgGluAlaGlyThrGluThrGlyGlyGluAspAspLySgGluGlu 398
DB 662 AGACAGGTATGAGGACAGGAGACTGAACGGGGGAGAAACATATATTAAGAGAA 721
QY 398 GGuLyLeLyAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProI 418
DB 722 AGAGAAACAGATGAATCTTGAGCTCTCTGAAGCAATTTCTGGTCAACACCAAT 781
QY 418 eLySMeLyProAsnIleGluProProGluAsnValGluTrSerGlyAlaGluAlaSe 438
DB 782 AAAGATGAGCCAAATATTGAACCTCTGAGAAATGTGAGTGAGTGCTGAGAGCTC 841
QY 438 rMetPheArgValLeuIleGlyThrTyTyTyAspAspPheCysAlaIleAlaArgLeuI 458
DB 842 AATGTTAGAGTCTCTGACTGCACTTACATGACAAATTTCTGCTCAATGAGTTAAT 901
QY 458 eGlyThrLySgThrCysArgGlnValTyGlu-PheArgValLySgLySgSerIleIleA 478
DB 902 CGGAGCCAAACATGTAGACAGGTATGATTATAGTCAAGTCAAGATCTAGCATATAG 961
QY 478 lAProAlaProAlaGluAspValAspThrProProArgLySgLySg 493
DB 962 CTCAGAGTCCCGCTGAGATGTGATCTCTCCAAAGAAAGAAAGA 1008

RESULT 15
US-11-266-748A-361225/c
/ Sequence 361225, Application US/11266748A
/ Publication No. US20060134663A1
/ GENERAL INFORMATION:
/ APPLICANT: Hartlin, Paul
/ APPLICANT: Mulligan, Karl
/ TITLE OF INVENTION: Transcription Microarray Technology and
/ FILE REFERENCE: 55815-0102 (319189)
/ CURRENT APPLICATION NUMBER: US/11/266,748A
/ CURRENT FILING DATE: 2005-11-03
/ PRIOR APPLICATION NUMBER: EP 04105479.2
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105482.6
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105483.4
/ PRIOR FILING DATE: 2004-11-03
```

```

1 PRIOR APPLICATION NUMBER: EP 04105507,
2 PRIOR FILING DATE: 2004-11-03
3 PRIOR APPLICATION NUMBER: EP 04105485.9
4 PRIOR FILING DATE: 2004-11-03
5 PRIOR APPLICATION NUMBER: EP 04105484.2
6 PRIOR FILING DATE: 2004-11-03
7 PRIOR APPLICATION NUMBER: US 60/662,276
8 PRIOR FILING DATE: 2005-03-14
9 PRIOR APPLICATION NUMBER: US 60/700,293
10 PRIOR FILING DATE: 2005-07-18
11 NUMBER OF SEQ ID NOS: 483996
12 SOFTWARE: PatentIn version 3.3
13 SEQ ID NO 361225
14
15 LENGTH: 904
16
17 TYPE: DNA
18
19 ORGANISM: Homo Sapiens
20
21 FEATURES:
22
23 NAME/KEY: misc feature
24 LOCATION: (30)..(30)
25 OTHER INFORMATION: n is a, c, g, or t
26
27 FEATURES:
28
29 NAME/KEY: misc feature
30 LOCATION: (134)..(134)
31 OTHER INFORMATION: n is a, c, g, or t
32
33 FEATURES:
34
35 NAME/KEY: misc feature
36 LOCATION: (139)..(139)
37 OTHER INFORMATION: n is a, c, g, or t
38
39 FEATURES:
40
41 NAME/KEY: misc_feature
42 LOCATION: (205)..(205)
43 OTHER INFORMATION: n is a, c, g, or t
44
45 FEATURES:
46
47 NAME/KEY: misc feature
48 LOCATION: (240)..(240)
49 OTHER INFORMATION: n is a, c, g, or t
50
51 US-11-266-748A-361225

```

THIS PAGE LEFT BLANK



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 13, 2006, 13:11:01 ; Search time 10481.6 Seconds  
(without alignments)  
6396.808 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGQCKSEKSEKGVCMKRRVK.....KGNKRPANHSVNPYCAKV 699

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abs/ABSSWB.spool/US10773302/runat.11082006.140223.16618/app.query.fasta\_1  
-DB=GenEmbl -QFMT=fastap -SUPPLX=p2n.rge -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human0.cdi -LIST=45  
-DOCLIGN=200 -THR.SCOR=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs804  
-USER=US10773302 @CGN 1.1.13924 @runat.11082006.140223.16618 -NCPU=6 -ICPU=3  
-NO MMAP -NNG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb ENV:\*  
2: gb PAT:\*  
3: gb PH:\*  
4: gb PL:\*  
5: gb PR:\*  
6: gb RO:\*  
7: gb STB:\*  
8: gb SY:\*  
9: gb UN:\*  
10: gb VL:\*  
11: gb OV:\*  
12: gb HEG:\*  
13: gb IN:\*  
14: gb OM:\*  
15: gb DA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	2511	2	CQ714720 Sequence
2	3808	100.0	2600	2	A56329 Sequence 1
3	3808	100.0	2600	2	AR473734 Sequence

4	3808	100.0	2600	2	AX333675	AX333675 Sequence
5	3808	100.0	2600	5	HSU61145	U61145 Human enhan
6	3803	99.9	2619	5	AB179385	AB179385 Macaca fa
7	3802	99.8	2558	7	BD186247	BD186247 STAYC act
8	3802	99.8	2576	2	AX821950	AX821950 Sequence
9	3802	99.8	2576	2	HS2ESTENH	X95653 H.sapiens m
10	3787.5	99.5	2655	5	BC010858	BC010858 Homo sapi
11	3781	99.3	2512	2	DD208658	DD208658 Homo sapi
12	3756	98.6	2574	6	BC003772	BC003772 Mus muscu
13	3756	98.6	2595	6	BC016391	BC016391 Mus muscu
14	3756	98.6	2600	6	AX220174	AX220174 Mus muscu
15	3750	98.5	2653	6	MMU52951	U52951 Mus muscu
16	3679	96.6	2617	6	BC079538	BC079538 Mus muscu
17	3586	94.2	2631	11	CR855647	CR855647 Xenopus t
18	3569.5	93.7	2476	2	AR18662	AR18662 Sequence
19	3569.5	93.7	2476	2	AX747667	AX747667 Sequence
20	3569.5	93.7	2476	5	AK092676	AK092676 Homo sapi
21	3566	93.6	2658	11	BC084193	BC084193 Xenopus 1
22	3550	93.2	2772	11	AF351126	AF351126 Xenopus 1
23	3540.5	93.0	3641	5	AB208895	AB208895 Homo sapi
24	3515	92.3	2636	11	BC097526	BC097526 Xenopus 1
25	3455.5	90.7	2546	5	AB168941	AB168941 Macaca fa
26	3138	82.4	3394	11	AB195561	AB195561 Oryzias 1
27	3119	81.9	2666	5	HSU52965	U52965 Human putat
28	2863	75.2	1960	2	BD186308	BD186308 STAT6 act
29	2625.5	68.9	4323	11	TN1333461	AJ313461 Tetracodon
30	2389.5	62.7	2253	6	AF483490	AF483490 Mus muscu
31	2389.5	62.7	2253	6	AF483491	AF483491 Mus muscu
32	2389.5	62.7	3970	6	MMU60453	U60453 Mus muscu
33	2389.5	62.7	4179	6	BC007135	BC007135 Mus muscu
34	2389.5	62.7	4185	6	AB004817	AB004817 Mus muscu
35	2386.5	62.7	2244	8	BT009782	BT009782 Homo sapi
36	2386.5	62.7	2244	8	AY889044	AY889044 Synthetic
37	2386.5	62.7	2594	2	CQ715926	CQ715926 Sequence
38	2386.5	62.7	4606	5	AB002386	AB002386 Homo sapi
39	2386.5	62.7	4640	5	BC015882	BC015882 Homo sapi
40	2384.5	62.6	2594	2	AX821949	AX821949 Sequence
41	2384.5	62.6	2594	5	HSU50315	U50315 Human enhan
42	2364	62.1	4639	5	HSMB0470	AL833159 Homo sapi
43	2362.5	62.0	4147	6	AK129004	AK129004 Mus muscu
44	2306	60.6	2749	5	AB004818	AB004818 Homo sapi
45	2082	54.7	1889	11	BC095094	BC095094 Danio rer

ALIGNMENTS

RESULT 1  
CQ714720  
LOCUS CQ714720 654 bp DNA 2511 bp linear PAT 03-FEB-2004  
DEFINITION Sequence CQ714720  
VERSION CQ714720  
KEYWORDS GI:42275577  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

REFERENCE

1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, B.W.

Kits, such as nucleic acid arrays, comprising a majority of  
human exons or transcripts, for detecting expression and other uses  
thereof

JOURNAL Patent: WO 02068579-A 654 06-SEP-2002;

FEATURES

source location/qualifiers

ORIGIN

1..2511  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

Alignment Scores:

Pred. No.: 0 Length: 2511  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0

US-10-773-302-2\_copy\_1\_699 (1-699) x CQ74720 (1-2511)

QY 1 MetGluGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB 10 ATGGGCCAGACTGGGAGAGAAATCTGAGAGGAGCAAGTTGTTGGCGGAGCGCTGTA 69  
QY 21 SerGluTrpMetCysArgLysGlnLysLysArgPheArgAlaAspGluValLysSer 40  
DB 70 TCAGAGTACATGCCGACTGACAGACTCAAGAGGTTCAACAGAGCTGATGAAGTAAAGGT 129  
QY 41 MetPheSerSerAsnArgGlnLysLysLeuGluLysGlnLysLysLysLysLysLys 60  
DB 130 ATGTTAGTTCACATCGTCAGAAAATTTGGAAAGAACGAAATCTTAAACAGAAATGG 189  
QY 61 LysGlnArgArgLysGlnProValHisLeuThrSerValSerSerLeuArgGlyThr 80  
DB 190 AAACGCGAAGGATACAGCTGTGTCATCTGACTTCTGAGCTCATTTGCGGAGACT 249  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 250 AGGAGGTGTCGGTACCACTGATCTTGGATTTTCCACACAAAGTCATCCCATTAAGACT 309  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnAspPhe 120  
DB 310 CTGATGAGAGTGTCTCAGTACCCATTAATGTAATCTTGCTCCCTACACAGAAATTT 369  
QY 121 MetValGluAspGlnThrValLeuHisAsnIleProLysMetGlyAspGluValLeuAsp 140  
DB 370 ATGTGTGAGAGATGAAACTGTTTACATTAACATCTCTTAATGAGAGATGAAGTTTAT 429  
QY 141 GlnAspGlyThrPheIleGluLysLeuLysAsnLysAspGlyLysValHisGlyAsp 160  
DB 430 CAGAGTGTACTTTCATTGAAGAACTTAATTAATAATTTATGATGGGAAAGTACACGGGAT 489  
QY 161 ArgGluCysGlyPheIleAsnAspGluLysPheValGluLeuValAsnAlaLeuGlyGln 180  
DB 490 AGAGATGTGGGTATTAATGATGAATTTTGTGAGGTGGTAAAGCCCTTGCTGTA 549  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluLysGluLysGln 200  
DB 550 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 610 AAAGATCTGAGAGATCACCGAGATGATTAAGAAACCCGCCACCTCGGAAATTTCTCT 669  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLysLeu 240  
DB 670 GATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGGACACGAGAAAGCTA 729  
QY 241 LysGluLysLysLysLysLysLysGlnLysGlnLysProGlyAlaLeuProProGluCys 260  
DB 730 AAGGAAAAATTAAGAACTCACCGAACAGAGCTCCAGGCGCACTTCTCTGTAAGT 789  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnLysSerLeuHisSer 280  
DB 790 ACCCCCAATATGATGAGCAAAATGCTAAATGTTTACAGAGAGACCAAGCTTACATCC 849  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysLysLysLysLysLysLysLysLys 300  
DB 850 TTTCTAGAGCTTTTCTGAGGCGATGTTTAAATGACCTGCTTCTCATCTCTTTTAT 909  
QY 301 AlaThrProAsnThrLysLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 910 GCAACACCCACATTAATTAAGCGAAGAACACAGAAACAGCTCTAGACACAAACCTTGT 969

QY 321 GlyProGlnCysTrpGlnHisLeuGluGlyAlaLysGlnPheAlaAlaLeuThrAla 340  
DB 970 GAGCAACAGGTTAACGACATTTGGAGGAGCAAGAGATTTCTGCTGCTCACCGCT 1029  
QY 341 GlnArgLysLeuThrProProLysArgProGlyGlyLysArgArgGlyArgLeuProAsn 360  
DB 1030 GAGCGGATTAAGACCCCAACCAACGTCAGAGAGCCGCGAGAAAGAGAGCGGCTTCCCAT 1089  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1090 AACAGTAGAGCCCGACACCCCAACCATTAATGTGTGAATCAAAAGATATACAGACGT 1149  
QY 381 AspArgGlnLysGlyThrGlnLysLysLysLysLysLysLysLysLysLysLysLys 400  
DB 1150 GATAGGAGACAGGAGCTGAACCGGGGAGAGAACATATTAAGAAAGAAAGAGAG 1209  
QY 401 LysAspGlnThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1210 AAGATGAACTTTCGAGCTCTCTGAAAGCAAAATTCGAGTCAAAACCAATTAAGATG 1269  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluLysMetPhe 440  
DB 1270 AAGCCAAATATTGAACCTCTGAGAAATGAGAGTGAAGTGTGCTGAAGCTCATGTTT 1329  
QY 441 ArgValLeuLysGlyThrLysLysLysLysLysLysLysLysLysLysLysLysLys 460  
DB 1330 AGAGTCTCATGTCAGCTTACTATGACAAATTTCTGTCACATGCTAGATTAATTTGGACC 1389  
QY 461 LysThrCysArgGlnValLysGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1390 AAAACATGTACAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1449  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysLysLysLysLysLysLys 500  
DB 1450 CCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1509  
QY 501 AlaHisCysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 520  
DB 1510 GCACATGCGAAGAAAGATACAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1569  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1570 CACCTGTGATCATCCACGCGAGCTTGTGACAGTTGCGTCCCTTGTGTGATGACAA 1629  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnAspPheProGlyCys 560  
DB 1630 AATTTTGTAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1689  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLysAlaValArgGluCys 580  
DB 1690 CGTGCAGAGCAGCTGACACCAACAGAGGCGCGTCACTGCTGCTGCTGCTGCTGCTGCT 1749  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1750 GACCTTACCTCTGCTCTTACTTGTGAGCGGCTGACCATTTGGACAGTAAATGTGTCC 1809  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuLeuAlaProSerAsp 620  
DB 1810 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGATCTATTTGCTGGGACCACTGAC 1869  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1870 GTGGCAGCGTGGGAGATTTTATTAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 1929  
QY 641 TyrCysGluGluLysLeuSerGlnAspGlyAlaAspArgArgGlyLysValLysAspLys 660  
DB 1930 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGGAAAGTATGATTA 1989  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
DB 1990 TACATGTGAGCTTCTGTTCAACTTGAACAATGATTTTGTGTGATGCAACCCGGAAG 2049  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTrpLysAlaLysVal 699

Db 2050 GGTAACAAATTCGTTTGCATATCGTAATCCAACTGATGCAAAAGTT 2106

RESULT 2  
LOCUS A58329 2600 bp DNA linear PAT 05-MAR-1998  
DEFINITION Sequence 1 from Patent WO9635784.  
ACCESSION A58329  
VERSION A58329.1 GI:3713991  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homosapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

REFERENCE  
1 Jenuwein, T. and Laible, G.  
AUTHORS CHROMATIN-REGULATOR GENES  
TITLE Patent: WO 9635784-A 1 14-NOV-1996;  
JOURNAL BOEHRINGER INGELHEIM INT. (DE)  
COMMENT Other publication DB 19516776 961114.  
FEATURES  
Location/Qualifiers  
source 1..2600  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/cell\_type="B-CELL"  
1..89  
90..2330  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAA03482.1"  
/db\_xref="GI:3713992"  
/translation="MGQTKSKSEKQPCVMKRVKSEYMLRLQLKRRFRADEVKSMSS  
NROKILBERTELINDEKORRLOPHILTSVSLRGTRECSYSDIDPTQVILPLTLN  
AAVAPIMYSPLQONFMVEDEVIAHNI PYMGDEVLPDQGFIELIKNDGKHGD  
RECGFINDEIFELYNALGOYNDDDDDGDDPERERKODLEHRRDKSRPRKX  
PSDKIFEAISSEMPDKGTABELKEXKYLTLQQLFGALPRECTPIIDPMKSVQRE  
SLSHFTILFCRCFKTDCFLPFHAFENTYRKKTETALMDNRCQPCYOHBEKKEF  
AAALTAERI KTPPKRGRGRGRGLPNNSRSTPTINVLSDTDSDEAGTERGEN  
NDKEBERKDETRSSSEANSRQPTIKKPNLSEPVNWSGAESMFRVILGYTDN  
FCALRLIGTKRCQVREFRVSESIILAPAPADVDTPPKKKRKHRLMAHAKRIOL  
KKDSSNVNVPDHPROPDSDSCPCVIANPCEKPCOCSSECONPFGRCACAC  
NTKQCPCTLAARECPDCLITCGAADHDSKVKSCNKSIOGSGKKHLLAPSDVAGH  
GIFIKDVPQKNEFISEYCGEIIISDEADRKGVYDKICSELPFNINDFVVDATKGN  
KIRFANHSVNPNCYAKVMVWNGDHRIIGI FAKRALQITGBELFPDYRYSGADMLKVVGIE  
REMEIP"

3'UTR  
ORIGIN 2331..2600

Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x A58329 (1-2600)

QY 1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysThrArgLysArgValys 20  
DB 90 ATGGGCGACACTGGGAGAAATCTGAGAAAGGACCAAGTTGTTGGCGAAGCGTGTAAAA 149  
QY 21 SerGlyTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 150 TCAGAGTCATCGCATGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAGAGTAAGAGT 209  
QY 41 MetPheSerSerAsnArgGlnLysLysLeuGluArgThrGlnLysLeuAsnGlnLysTrp 60  
DB 210 ATGTTAAGTCCAACTGTCAGAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATGG 269  
QY 61 LysGlnArgArgLysGlnProValHisLysLeuThrSerValSerSerLeuArgLysThr 80

Db 270 AAACGCGAAGGATACACCTGTGACATCCGACTTCTGTGAGCTCATTTGGCGGACT 329  
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValLysProLeuLysThr 100  
DB 330 AGGAGTGTTCGTGTCGACGACGATGATGATTTTCCAAACAAAGTCATCCATTAAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProLysMetTyrSerTrpSerProLeuGlnAsnPhe 120  
DB 390 CTGAATGCAAGTGTCTTCATGACCACTAATGATTTTGGTCTCCCTACGACGAATTTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnLysProTyrMetGlyAspGluValLeuAsp 140  
DB 450 ATGCTGCAAGATGAAACGTTTACATTAACATCTCTTATATGGAGATGAAAGTTTAAAT 509  
QY 141 GlnAspGlyThrPheLysGlnLysLeuLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 510 CAGATGTGATCTTTCATGAAAGACTAATTAATAAATTTATGATGGGAAAGTACACGGGAT 569  
QY 161 ArgGlyCysGlyPheLysAsnAspGlnLysPheValGluLeuValAsnAlaLeuGlyGln 180  
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGGTAATGCTTGCTCA 629  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
DB 630 TTTAATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 201 LysAspLeuGluAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 690 AAAGATCTGAGGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 749  
QY 221 AspLysLysLysPheGlnAlaLysSerSerPheProAspLysGlyThrAlaGlnLysLeu 240  
DB 750 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGACACGACGAACACTA 809  
QY 241 LysGlnLysTyrLysGlnLeuThrGlnGlnLysLeuProGlyAlaLeuProProGlyLys 260  
DB 810 AAGGAAATATTAAGAACTACACCGAACGACGCTCCAGGCGCATCTCTCTGAATGT 869  
QY 261 ThrProAsnLysAspGlyProAsnAlaLysSerValGlnArgGlnLysLeuHisSer 280  
DB 870 ACCCCCAATATTAAGATGACCAATAGCTTAATCTGTCGAGAGAACAAAGCTTACATCC 929  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 930 TTTCAATAGCTTTCTGTAGGCGATGTTTAAATATGATCTCTCTCAATCCCTTTTAT 989  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB 990 GCAACACCCACACTTATATAGCGAAGAACACGAAACAGCTCTAGACACAAACCTTGT 1049  
QY 321 GlyProGlnCysTyrGlnHisLeuGlnGlyAlaLysGlnPheAlaAlaLeuThrAla 340  
DB 1050 GGACCAACAGTGTTCACGCAATTTGAGAGAACAAAGAGTGTTCGCTGCTCAACGCT 1109  
QY 341 GlnArgGlnLysLysThrProProLysArgProGlyGlyArgArgGlyLysLeuProAsn 360  
DB 1110 GAGCGGATTAAGACCCCAACCAAAAGTCCAGAGAGCGCGCAAGAGGACGACTCCCAAT 1169  
QY 361 AsnSerSerArgProSerSerThrProThrLysAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1170 AACAGTACAGAGCCCAAGACCCCAACATTAATGTCTGAAATCAAGAGATACAGACGT 1229  
QY 381 AspArgGlnAlaGlyThrGlnThrGlyGlyGlnAsnAsnAspLysGlnGlnGlnLys 400  
DB 1230 GATAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1289  
QY 401 LysAspGlnThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProLysMet 420  
DB 1290 AAAGATGAAATTTGAGACTCTCTGAAAGCAATTTCTGGGTCTCAACCAATTAAGATG 1349  
QY 421 LysProAsnLysGlnProProGlnAsnValGlnLysSerGlyAlaGlnLysSerMetPhe 440

Db 1350 AAGCCAAATATGAACTCTGAGAAATGTGAGTGTGCTGAAGCTCAATGTTT 1409  
Qy 441 ArgValLeuIleGlyThrTyrTyrAspAsnPhcCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1410 AGAGTCTCTCATTTGGACCTTACTATGACAAATTTCTGTGCACTTGTAGTTAAATTTGGACC 1469  
Qy 461 LysThrCysArgGlnValTyrGluPheArgValIlysgIuserSerIleIleAlaProAla 480  
Db 1470 AAAACATGTAGACAGTGTATGAGTTTGTAGTCAAAAGATCTACATCATAGTCCAGCT 1529  
Qy 481 ProIleGluAspValAspThrProBroArgLysLysValAspGlyValAspLeuThrAla 500  
Db 1530 CCCCCTGAGAGTGTGATATCTCTCAAGAAAAGAGAGAAACCCGGTGTGGCT 1589  
Qy 501 AlAlaIscCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
Db 1590 GCACACTGCAGAAAAGATACACTGAAAAAGACGGCTCTCTAACCAATGTTTACACATAT 1649  
Qy 521 GlnProCysAspPheIAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1650 CAACCCCTGTGATCATCCACGGCAGCCTGTGACAGTTGCTGCCCTGTGTATAGCAAA 1709  
Qy 541 AsnPhcCysGlyLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
Db 1710 AATTTTGTGAAAAGTTTGTCAATGTATGTTCAAGTGTCAAAACCGCTTTCGGGATGC 1769  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
Db 1770 CGCTGCAGAAAGCAGACTGCACACCAAGCAGATGCCGTGCTACCTGCTGCAGAGATGT 1829  
Qy 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisIlePheAspSerLysAsnValSer 600  
Db 1830 GACCTGACCTCTGCTTACTTGTGTGAGCCCTGACCATTTGGGACAGTAAAAATGTGTCC 1889  
Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleuLeuLeuAlaProSerAsp 620  
Db 1890 TGCAGAACTGAGATATTCAGCGGGCTCCAAAAGACATATATGCTGCAACATCTGAC 1949  
Qy 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 1950 GTGCGAGCTGGGGATTTTATTCAAAGATCTCTGCGAGAAAAGATTCATCTCAAA 2009  
Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValIleThrAspLys 660  
Db 2010 TACTGTGAGAGATATTTCTCAAGATAGAGCTGACAGAGCGGAAAGTGTATGATATA 2069  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrAspLys 680  
Db 2070 TACATGTGAGCTTTCTGTCTCAACTTGAACATGATTTTGTGTGATGCAACCGCAAG 2129  
Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
Db 2130 GGTAAACAAATTCGTTTTCGAAATCATTCGTTAAATCCAACTGCTAAGCAAAAGTT 2186

RESULT 3  
AR473734 2600 bp DNA linear PAT 20-FEB-2004  
LOCUS AR473734  
DEFINITION Sequence 1 from patent US 6689583.  
ACCESSION AR473734  
VERSION AR473734.1 GI:42712196  
KEYWORDS  
SOURCE  
ORGANISM  
Unknown.  
Unclassified.  
REFERENCE  
1 (bases 1 to 2600)  
AUTHORS Jernwein,T., Laidle,G., O'Carroll,D., Eisenhaber,F. and Rea,S.  
TITLE Chromatin regulator genes  
JOURNAL Patent: US 6689583-A 1 10-FEB-2004;  
Boehringer Ingelheim International GmbH;;  
DEX;  
FEATURES  
SOURCE location/Qualifiers  
1..2600  
/organism="unknown"

ORIGIN /mol\_type="genomic DNA"  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0  
US-10-773-302-2\_copy\_1\_699 (1-699) x AR473734 (1-2600)  
Qy 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTyrPheArgValLys 20  
Db 90 ATGGGCCAGACTGGAGAAATCTGAGAAAGGACCAATTTGTGGCGGAAACCGTAA 149  
Qy 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
Db 150 TCAGAGTACATGCAGCTGAGACAGCTCAAGAGGTTCAAGAGCTGATGAAATGAAGT 209  
Qy 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnLys 60  
Db 210 ATGTTTATGTTCCATGCTCAGAAATTTTGGAAAGAACGAAATCTTAAACAAAGATGG 269  
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 270 AAACAGCGAAGATACAGCTGTGCACATCTGCTTGTGAGCTCATTTGCGGGACT 329  
Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 330 AGGAGGTGTCGGAGCAGTGAAGTGAATTTTCCACACAGTATCCATTTAAAGACT 389  
Qy 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrPheProLeuGlnAsnPhc 120  
Db 390 CTGAATCAGTGTGTTCAATACCATATATGATTTCTGTGCTCCCTACAGCAGAAATTTT 449  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProCysTyrMetGlyAspGluValLeuAsp 140  
Db 450 ATGCTGCAAGATGAAATCGTTTATCATTAACATTCCTATATGGAGATGAAGTTTATGAT 509  
Qy 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
Db 510 CAGAGTGTACTTTCATTTGAAGACCTAATTAATAATTAATGATGGAAAGTACCGGGAT 569  
Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 570 AGAAGATGTGGTTTATTAATGATGAATTTTGTGAGTTGGTGAATGCCCTTGTCAA 629  
Qy 181 TyrAsnAspAspAspAspAspAspArgLysAspAspProGluLysGluGluLysGln 200  
Db 630 TATATATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
Qy 201 LysAspLeuGluAspPheIAspAspAspLysGluSerArgProProArgLysPheProSer 220  
Db 690 AAAGATCTGAGGATCACCGAGATGATTAAGAAAGCCGCCACCTCCGAAATTTCTTCT 749  
Qy 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
Db 750 GATTAATTTTGAAGCATTTCTCTCAATGTTTCAAGTAAAGGACACAGCAAGAAACTA 809  
Qy 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGluValAlaLeuProProGluCys 260  
Db 810 AAGGAAAAATATAAGAACTCACCGAAGCAGAGCTCCAGCGCATCTTCCCTGAATGT 869  
Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
Db 870 ACCCCCAACATAGATGACCAAAATGCTAAATCTGTCAAGAGAGCAAAAGCTTACATCC 929  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisIleProPheHis 300  
Db 930 TTTCATAGCGCTTTCTGTAGGCGATGTTTAAATATATGATGATGATGATGATGATGAT 989

QY 301 AATThProAsnThrTyrIysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 DB 990 GGAACCCCACTTATTAAGCGAAGAACAGAAACAGCTTACAGCAACAACTTCT 1049  
 QY 321 GlyProGlnCysTyrGlnHisLeuGluGlyAlaLysGluPheAlaAlaLeuThrAla 340  
 DB 1050 GGACCAACAGTGTATCAGACATTTGGAGGAGCAAGAGATTGCTCTCTCACCGCT 1109  
 QY 341 GUAAGGLeuGlyThrProProLysArgProGlyValArgArgArgGlyValArgLeuProAsn 360  
 DB 1110 GAGCGATTAAGACCCCAACAAACCTCCAGAGCGCGAGAGAGAGCGCTTCCCAT 1169  
 QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 DB 1170 AACAGTAGCAGCGCCAGCAACCCCAACATTAATGCTCTGATCAAGAGATACAGACAGT 1229  
 QY 381 AAPATGGLuAlaGlyThrGluThrGlyGlyLysAsnAspLysGluGluGluLys 400  
 DB 1230 GATAGGAGAGCAGGAGCTGAACCGGGGAGAGAACATGATTAAGAGAGAGAGAG 1289  
 QY 401 LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 DB 1290 AAGAGTGAACCTTCAGCTCCCTGTAAGCAAAATCTCGGTGTCAACACCAATTAAGATG 1349  
 QY 421 LysProAsnIleGluProProGluAsnValGluTyrSerGlyValAlaLysSerMetPhe 440  
 DB 1350 AAGCCAAATATTGAACCTCTGAGATGTGAGGTGAGTGTGCTGAGACCTCAATGTTT 1409  
 QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 DB 1410 AAGATCTCATTTGGCAGCTTACTATGACATTTCTGTGCAATTTGTAGGTTAATGGAGCC 1469  
 QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 DB 1470 AAAACATGTAGACAGGTGTATAGATTAGACCAAGATCTGACATCTACCTCAGCT 1529  
 QY 481 ProAlaGluAspValAspThrProProArgLysLysArgLysHisArgLeuThrAla 500  
 DB 1530 CCCCCTGAGATGTGATCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589  
 QY 501 AlaHisCysArgLysIleGlnLeuLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 DB 1590 GCACACTGCAGAAAGATACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1649  
 QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 DB 1650 CAACCTGTGATCAATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1709  
 QY 541 AsnPheCysGluLysPheCysGlnCysSerSerSerGluCysGlnAsnArgPheProGlyCys 560  
 DB 1710 AATTTTGTGAAGATTGTTCATATGTAGTTCAGAGTTCAGAAACCGCTTCCGGAGATGC 1769  
 QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys 580  
 DB 1770 CGCTCAAGAGCAGATGAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1829  
 QY 581 AspProAspLysCysLeuThrCysGlyAlaAlaAspHisArgPheSerLysAsnValSer 600  
 DB 1830 GACCTGACCTCTGTCTTACTGTGTGAGAGCGCTGACCAATGGAGAGAGAGAGAGATGTGCC 1889  
 QY 601 CysLysAsnCysSerIleGluArgGlySerLysLysHisLeuLeuLeuAlaProSerAsp 620  
 DB 1890 TGCAGAACTGCAGATTCAGCGGAGCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1949  
 QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 DB 1950 GGGGAGGCTGGGGAGATTTTATCAAAAGATCTGTGCAAGAGAGAGAGAGATTCACAGAA 2009  
 QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAGAGAGAGAGATGTATATAA 2069  
 QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680

DB 2070 TACATGTGAGCTTTCTGTTCACCTGAACATGATTTTGTGTGATGCAACCGGAG 2129  
 QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 DB 2130 GGTAACAAATTCGTTTGGCAATCATTCGGTAATCCAAATCGCTATGCAAAAGTT 2186  
 RESULT 4  
 AX333675  
 LOCUS AX333675 2600 bp DNA linear PAT 09-JAN-2002  
 DEFINITION Sequence 4184 from Patent WO0194629.  
 ACCESSION AX333675  
 VERSION AX333675.1 GI:18124394  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 REFERENCE  
 1  
 AUTHORS Young, P.B., Augustus, M., Carter, K.C., Edner, R., Endress, G.,  
 Horrigan, S., Soppet, D.R. and Weaver, Z.  
 TITLE Cancer gene determination and therapeutic screening using signature  
 gene sets  
 JOURNAL Patent: WO 0194629-A 4184 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 FEATURES  
 source  
 location/qualifiers  
 1..2600  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 2600  
 Score: 3808.00 Matches: 699  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 2 Gaps: 0  
 US-10-773-302-2\_COPY\_1\_699 (1-699) x AX333675 (1-2600)  
 QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
 DB 90 ATGGGCCAGCTGGAGAAATCTGAGAGGAGACCAAGTTGTGGCGAGACGTTAA 149  
 QY 21 SerGluTyrMetCysArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
 DB 150 TCAGAGTACATGCCATGACAGACAGCTCAAGAGGTTCAAGACAGCTGAGAGATTAAGAT 209  
 QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
 DB 210 ATGTTTAGTTCCAATCGTCAGAAATTTGGAAAGAGAGAAATCTTAAACCAAGATGG 269  
 QY 61 LysGlnArgArgGlyIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
 DB 270 AAACAGCGAAGATTAAGCGCTGTGCACATCTGTGACCTCATTCGCGGAGACT 329  
 QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
 DB 330 AGGAGATGTTGGGAGACAGAGTGAATTTTCCACACAGATCCCATTAAGACT 389  
 QY 101 LeuAsnAlaValAlaSerValProIleTyrSerTyrSerProLeuGlnAsnPhe 120  
 DB 390 CTGAATGAGATGCTTCAGTACCATATGATATCTTGCTCCCTTACAGACAGAAATTT 449  
 QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
 DB 450 ATGGTGAAGATGAACCTGTTTACATTAACATTCCTTAATAGGAGATGAAGATTAGAT 509  
 QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160

Db 510 CAGGATGGTACTTTCATTGAAGAACTAATAAAAAATTATGATGGAAAAGTACACGGGAGT 569  
 Qy 161 ArgGluCyseGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
 Db 570 AGAAGAAATGGGGTTTATTAATGATGAAATTTTGTGGAGTTGGTGAATGCCCTTGGCTCA 629  
 Qy 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluGln 200  
 Db 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
 Qy 201 LysAspLeuGluAspPheIleArgAspAspGlyLeuSerArgProProArgGlyPheProSer 220  
 Db 690 AAAAGATCTGGAGATGATCAACGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTCTT 749  
 Qy 221 AspLysIlePheGluAlaIleSerSerMetPheProAspGlyGlyThrAlaGluGluLeu 240  
 Db 750 GATAAATTTTGTGAAGCCATTTCTCAATGTTTCCAGTTAAGGCGACAGCAAGAAAGAACTA 809  
 Qy 241 LysGluGlyTyrGlyGluLeuThrGluGluGlnGluLeuProGluValAlaLeuProGluGlu 260  
 Db 810 AAGGAAAAATATAAAGAACTCACCGAAACAGCAGCTCCACGGCGCACTTCTCTGAATGT 869  
 Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
 Db 870 ACCCCCAACATAGATGAGCAAAATGCTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 929  
 Qy 281 PheHisThrLeuPheCyseArgArgCysePheLysTyrAspCysePheLeuHisProPheHis 300  
 Db 930 TTTCTACAGCTTTTCTGTAGCGAGATGTTTAAATAGATGATGATGATGATGATGATGATGAT 989  
 Qy 301 AlaThrProAsnThrTyrTyrAspGlyAspAsnThrGluThrAlaAlaAspAsnLysProCyse 320  
 Db 990 GCAACACCCCAACATTAATAGCGGAAGAAACAGAAACAGCTCTAGAACAAACCTTGT 1049  
 Qy 321 GlyProGlnCyseTyrGlnHisLeuGluGluValAlaLysGluPheAlaAlaAlaLeuThrAla 340  
 Db 1050 GGACCAAGGTGTACACGATTTGGAGGAGCAAAAGAGTTGCGTCTCTCACCGCT 1109  
 Qy 341 GluArgIleLysThrProProLysArgProGlyGlyArgArgGlyArgLeuProAsn 360  
 Db 1110 GAGCGGATTAAGAACCCCAACCAAAAGCTTCAGAGAGCCCGCAAGAGAGAGCGCTTCCCAAT 1169  
 Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 Db 1170 AACAGTAGCAGGCGCCAGCAGCCGCCACATTAATGTGCTGAATCAAGATACAGACAGT 1229  
 Qy 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluLys 400  
 Db 1230 GATAGGGAAGCAGGAGCTGAACCGGGGAGAGAAACAATGATTAAGAAAGAGAGAGAG 1289  
 Qy 401 LysAspGluThrSerSerSerSerSerGluIleAsnSerArgCyseGlnThrProIleLysMet 420  
 Db 1290 AAAAGTGAACCTTCAGAGCTCTCTGAAGCAAAATTTCCGTGTCAAAACCAATAAAGATG 1349  
 Qy 421 LysProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 Db 1350 AAGCCAAATATGAACCTCTGAGATGTGAGTGAAGTGGTGTGAAGCCCTCAATGTGTTT 1409  
 Qy 441 ArgValIleGluIleGlyThrTyrTyrAspAsnPheCyseAlaIleAlaArgLeuIleGlyThr 460  
 Db 1410 AGAGTCTCATGTCGACCTTACTATACATTTCTGTGTCATGTGTAAGTTAATGGAAACC 1469  
 Qy 461 LysThrCyseArgGluValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 Db 1470 AAAACATGTAGCAGGTGTATAGATTAGAGTCAAAGATCTAGCATCATAGCTCAGCT 1529  
 Qy 481 ProAlaGluAspValAspThrProProArgGlyLysLysValArgLysHisArgLeuThrAla 500  
 Db 1530 CCGCGTAGAGATGTGGATACTCTCTCCAAGGAAAAAGAGAGAAACACCGGTTGTGGGCT 1589  
 Qy 501 AlaHisCyseArgGlyValIleGluLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 Db 1590 GCACACTGCAGAAAGATACAGCTGAAAAAGACGCGCTCTTACCAAGTTTACAACTAT 1649

Qy 521 GlnProCyseAspHisProArgGlnProCyseAspSerSerCyseProCyseValIleAlaGln 540  
 Db 1650 CAACCTGTATCATCCACCGGAGCCCTTGAGACAGTTCCGCGCCCTGTGTGATGACAA 1709  
 Qy 541 AsnPheCyseGluLysPheCyseGlnCyseSerSerGluCyseGlnAsnArgPheProGlyCyse 560  
 Db 1710 AATTTTGTGAAAAAGTTTGTCAATGTAGTTCAAGAGTCAAAACGCTTTCCGGAGATGC 1769  
 Qy 561 ArgCyseLysAlaGlnCyseAsnThrLysGlnCyseProCyseTyrLeuAlaValArgGluCyse 580  
 Db 1770 CGCTGCAACACAGTGCACACCAACAGATGCCCGCTGACTACCTGGCTGCCAGAGTGT 1829  
 Qy 581 AspProAspLeuCyseLeuThrCyseGlyAlaAlaAspHisThrAspSerLysAsnValSer 600  
 Db 1830 GACCTGACCTCTGTCTTACTGTGAGCGCGCTGACCATTTGGACAGTAAATGTGTCC 1889  
 Qy 601 CyseLysAsnCyseSerIleGlnArgGlySerLysLysIleLeuLeuAlaProSerAsp 620  
 Db 1890 TGCAGAACTGCAAGTATTCAGCGGAGCTCCAAAGCATTTATGCTGGCAACATCTGAC 1949  
 Qy 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1950 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCTGTGCAAAAAATGAATTCATCTCAGAA 2009  
 Qy 641 TyrCyseGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGAGGAAAGTGTATGATAA 2069  
 Qy 661 TyrThrCyseSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
 Db 2070 TACATGTGCAAGCTTTCTGTCACTTGAACATGATTTTGTGTGATGACACCGGAG 2129  
 Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCyseTyrAlaLysVal 699  
 Db 2130 GTTACCAAAATGCTGTTTGCAAATCATTCGTAAATCAAACTGCTATGCAAAAGTT 2186

# RESULT 5

HSU61145  
 LOCUS 2600 bp mRNA linear PRI 11-JUN-1997  
 DEFINITION Human enhancer of zeste homolog 2 (EZH2) mRNA, complete cde.

VERSION 061145.1 GI:1575348  
 KEYWORDS

SOURCE Homo sapiens (human)  
 ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo

REFERENCE 1 (bases 1 to 2600)  
 Laible,G., Wolf,A., Dorn,R., Reuter,G., Nislow,C., Lebersorger,A.,  
 Popkin,D., Pillus,L. and Jenuwein,T.

TITLE Mammalian homologues of the Polycomb-group gene Enhancer of zeste  
 mediate gene silencing in Drosophila heterochromatin and at S.  
 cerevisiae telomeres

JOURNAL EMBO J. 16 (11), 3219-3232 (1997)  
 PUBMED 9214638

REFERENCE 2 (bases 1 to 2600)  
 Laible,G., Lebersorger,A. and Jenuwein,T.

AUTHORS Direct Submission  
 TITLE Submitted (18-JUN-1996) I.M.P., Dr. Bohrgasse 7, Vienna A-1030,  
 Austria

FEATURES  
 source Location/Qualifiers

1..2600  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_line="BWA-B"  
 /cell\_type="B-cell"  
 1..2600  
 /gene="EZH2"  
 90..2330  
 /gene="EZH2"

/note="Allele: E2H1; homolog of the Drosophila polycomb group protein enhancer of zeste encoded by GenBank accession number U00180"

/codon\_start=1

/product="enhancer of zeste homolog 2"

/protein\_id="AAC51520.1"

/db\_xref="gi:1575349"

/translation="MGQGGKSKGKGVCMRRVSEWRILQKLRPRADVKSMFSS  
NRKQIERTETLNQEMKORIQPHNIIITVSLSKRTCSYTSDDPFTQVITPLTAN  
AVASVPLIMWSPLQDFWEDVETLNIPTMGDEVLQDFLEELIKNDGKHGD  
RECGFINDEIVELVNALQYNDDDDDGDDPEREKKQDLBHRDNDKSRPRK  
PSDKIFEASISMPDKGTAEELKEKYEKELTEQLGALPPECTPIIDGNKSYOREQ  
SLHSFHTLCRCRCFYDCFLHPFATPTVTKRNTETLNDKPCPOCYHLEAKER  
AALTAERIKTPKRRGRRGRLPNNSRPSPTPIVTLKSGDDEAECTGEGN  
NDKREBEKDETSSSEANSRCQPIIKKRNIPBPENYBEGAEKSNFVITGTYDN  
FCALARLIGTKCRQVSEFRVSESIAPAPAEVDVTPRRKKRRLMAHCRKIQL  
KKDSSNHNVNYQPCDHPROPDSCPCVIAQNFCEKSCSECONFPCCRCACQ  
NTKQPCYLAVERCDPLCTGCADHDSKNVSKNCSIORSKGLHLAPSDVAGM  
GIFIKDPQKNEFISEYCGEIIISDEBARSKVYDKYKCSFLFNANDFVVDATKGN  
KIRPANHNVNCAKVMVNGDHRIGIPAKRALIOTGSELPFDIRYSQADLAKYVIGIE  
RMEELP"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 5

US-10-773-302-2\_COPY\_1\_699 (1-699) x HSU61145 (1-2600)

QY 1 MetGlyIntHglYlyLeuYseSergLulYsgLProValCyETPRARglYbArgYValYs 20  
DB 90 ATGGGCGACGACTGGGAGAAATCTGAGAGGAGCAGATTGTTGGCGAGAGGTATAAA 149  
QY 21 SerGluYrMetHargLeuArgLleuLylarGpHeaArgAlaAspGluValLysSer 40  
DB 150 TCAGAGTACATGCGACTGACGACCTCAAGAGTTCACACGACTGAGTAAAGTAAAGGT 209  
QY 41 MetPheSerSerAsnArgLulYsIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
DB 210 ATGTTTACATTCGATCGTCAAGAAATTTTGGAAAGACGGAATTTTAAACCAAGAAATGG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 270 AAACAGCGAAGGATACAGCTGTCGACATCTGACTTGCTGAGCTCATTTGGCGGAGCT 329  
QY 81 ArgGluCySerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 330 AGGAGTGTTCGGTACCACTGATGATTTTCCACACCAAGTCATCCCATTAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnLysAsn 120  
DB 390 CTGAATGCAATTGCTTCGATCCCAATGATTTTGTGCTCCCTCAAGCAAGAAATTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProYrMetGlyAspGluValLeuAsp 140  
DB 450 ATGGGGAAGATGAACCTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTATAT 509  
QY 141 GlnAspGlyThrPheIleGluGluLeuLleLysAsnYrAspGlyLysValHisGlyAsp 160  
DB 510 CAGATGTGATCTTTCATTTGAAGAACTAATTAATAATTTGATGGGAAAGTACACGGGAT 569  
QY 161 ArgGluCySgLYpHeIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluYs 180  
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGGGAAATGCCCTTGCTAA 629  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluLysGluLysGln 200  
DB 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689

QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 690 AAAGATCTGGAGGATCACCGAGATGATTAAGAAAGCCGCCACTCGGAAATTTCTTCT 749  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 750 GATAAATATTTTGAAGCATTCTCCATATGTTTCCAGATTAAGGACACGACGAAGACTA 809  
QY 241 LysGluYsYrYrYsGluLeuThrGluGlnGluLeuProGluYrAlaLeuProGluLys 260  
DB 810 AAGAAAAATTTTAAAGACTACCAAGAACGAGCTCCAGGCGCACTTCTTCGAAAGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 870 ACCCCCAATGATGACCAACCAATGCTTAATCTGTTCAAGAGAGCAAAAGCTTACACTCC 929  
QY 281 PheHisThrLeuPheCySarArgCySgPheLysYrAspCySgPheLeuHisProPheHis 300  
DB 930 TTTCAATGAGCTTTTCTGTAGGCGATGTTTAAATATGATGCTCTTCTCATCTTTTCT 989  
QY 301 AlaThrProAsnThrYrYrYsArgLysAsnThrGluThrAlaLeuAspAsnLysProCyS 320  
DB 990 GCAACACCCCAACACTTATAGCGGAGAACACAGAAACGCTCTAGACCAACAACTTGT 1049  
QY 321 GlnProGlnCySgYrGlnHisLeuGluGluValAlaYsGluPheAlaAlaLeuThrAla 340  
DB 1050 GGACCAAGTGTACAGCACTTGTGAGGAGCAAAAGATTTGCTGCTCTCACCGCT 1109  
QY 341 GlnArgIleYsThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360  
DB 1110 GAGCGGATTAAGACCCCAACCAACGTCACAGAGGCGGACAGAGGAGCGCTTCCCAT 1169  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1170 AACGATGACAGGCCACAGACCCCACTATATGTGCTGATCAAGAGATACAGACGT 1229  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluLys 400  
DB 1230 GATAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCySgIntnProIleYsMet 420  
DB 1290 AAAGATGAACCTTGAGCTCTCTGAGAGCAATTTCTGGTCTCAACCAACCAATTAAGATG 1349  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1350 AAGCCAAATATTTGAACCTCTGAGAGATGAGAGTGTGCTGAGAGCTTCATATGTTT 1409  
QY 441 ArgValLeuIleGlyThrYrYrYrAspAsnAspPheCySgAlaIleAlaArgLeuIleGlyThr 460  
DB 1410 AGAGTCTTCATTTGCACTTACTATGACAAATTTCTGTCCATTTGCTAGGTTAATGGGACC 1469  
QY 461 LysThrCySarGlnValYrGluPheArgValYsGluSerSerIleIleAlaProAla 480  
DB 1470 AAAACATGTACAGCGGTATGATGATTTAGAGTCAAAAGATTTACATCATGCTCCAGCT 1529  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysValYsArgLysHisArgLeuTrpAla 500  
DB 1530 CCCGCTGAGATGTGATATCTCTCCCAAGGAAAAAGGAAACACCGGTTTGGGCT 1589  
QY 501 AlaHisCySarArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValYrAsnThr 520  
DB 1590 GCACACTGCAAGAAAGATACAGCTGAAAAAGGACGCTCTCTTAACATGTTTACACTAT 1649  
QY 521 GlnProCySarAspHisProArgGlnProCySarAspSerCySgProCySValIleAlaGln 540  
DB 1650 CAACCTGTATCATTCACCGGACGCTTGTGACAGTTCTGTGCTTGTGTGATGACAA 1709  
QY 541 AsnAspCySgLYpPheCySgLYpCySgSerSerGluCySgLeuAsnArgPheProGlyCyS 560  
DB 1710 AATTTTGTGAAAAAGTTTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGAT 1769  
QY 561 ArgCySgLYpAlaGlnCySgAsnThrLysGlnCySgProCySgYrLeuAlaValArgGluCyS 580



[illegible]

distribution; clone distribution information can be found at:  
<http://www.nih.gov/yoken/genebank/>

Lab host: TOPIO  
Vector: pME18-FL3 (Acc.No. AB009864)  
R. Site1: DraIII (CACTGTGCG)  
R. Site2: DraIII (CACCATGGG)  
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraII sites of pME18-FL3. XhoI sites just outside the DraII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method. Libraries were made from:

Oocyte: cerebellum cortex  
ONPA: parietal lobe  
OCRA: temporal lobe right  
OFIA: frontal lobe left  
QOMA: medulla oblongata  
QBSA: brain stem  
OCRA: occipital lobe right  
QCSA: testis

Custom primers were used for 5' and 3'-end sequencing. The full-insert sequencing was done by primer-walking method using ABI DNA sequencer.

FEATURES	source																					
CDS	<p>Location/Qualifiers</p> <p>1..2619</p> <p>/organism="Macaca fascicularis"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9541"</p> <p>/clone="QCSA-18821"</p> <p>/sex="male"</p> <p>/clone_id="macaque cDNA library QCSA"</p> <p>/dev_stage="adult"</p> <p>93..2233</p> <p>/note="unnamed protein product; Homo sapiens enhancer of zeste homolog 2 (Drosophila) (EZH2), transcript variant 2, mRNA, RefSeq: NM_152998.1"</p> <p>/codon_start=1</p> <p>/protein_id="BAB02436.1"</p> <p>/db_xref="GI:67972188"</p> <p>/translation="MGQTGKSEKGPVCMWRKYSEYMR.LQLKRPRADEVKSMFSS/NROKLRTETLINOEQRRRIOPPHILITVSLSLRKEGSTSDLPPTQVIPLKTINAAVASVPIMYSMSPLQCNPMVEDVTVALNIPMDGRVLDDOGTFIELIKVDGVHKGRECFINDDELFEVLNALGCYTNDDEDDDDPREBREKQDLDRHDDESRPRKRPSDDIFSAISMPFDKGTABELKEKYELTEQLPGALPEPCTNTIGPNAKSQORECSLSHFHTLFCRCRCPGCRFLHPFAFTANAKRKNTETALDNKPCGPCYOHLBAKEAALTAARIKTPPPKPGRRGRGLJNNNSRSPSTPTIVNLBSKDSDSBAGETEGENNDKEERKDEBTSSSSBANSRCOPTIIMKPVIEPENVMWGAASMFRLVIGTYDNPCLAIRLGITKTCROYVEFRYKESIIAPAEVDNPPRKKKRHLMAMHCKRILOKDGSSSNHYNYIQPCDHPRDCSSCSCVLAONCEKRFCCOSSCONRPFRCACQNTKQPCPYLAIRECDPLCTGAGADHWDSGNVSKNCSIRGSKHILLAPSVAGKGIIFCPYONKERTISEYCGELISQDEADRGRKVDKMCSTFLNANDVFVDAIRKGGKIRFANNSVNPNCYAKTMVMVGDIRIGIIPAKRAIQGBELFYDYRSQDALKVYGLIHREMEIP"</p>																					
ORIGIN	<p>Alignment Scores:</p> <table border="1"> <thead> <tr> <th>Pred. No.:</th><th>Length:</th><th>Matches:</th></tr> </thead> <tbody> <tr> <td>Score:</td><td>0</td><td>2619</td></tr> <tr> <td>Percent Similarity:</td><td>3803.00</td><td>698</td></tr> <tr> <td>Percent Identity:</td><td>99.9%</td><td>Conservative: 0</td></tr> <tr> <td>Best Local Similarity:</td><td>99.9%</td><td>Mismatches: 1</td></tr> <tr> <td>Query Match:</td><td>99.9%</td><td>Indels: 0</td></tr> <tr> <td>Ds:</td><td>5</td><td>Gaps: 0</td></tr> </tbody> </table> <p>US-10-773-302-2_COPY_1_699 (1-699) x ABI79385 (1-2619)</p>	Pred. No.:	Length:	Matches:	Score:	0	2619	Percent Similarity:	3803.00	698	Percent Identity:	99.9%	Conservative: 0	Best Local Similarity:	99.9%	Mismatches: 1	Query Match:	99.9%	Indels: 0	Ds:	5	Gaps: 0
Pred. No.:	Length:	Matches:																				
Score:	0	2619																				
Percent Similarity:	3803.00	698																				
Percent Identity:	99.9%	Conservative: 0																				
Best Local Similarity:	99.9%	Mismatches: 1																				
Query Match:	99.9%	Indels: 0																				
Ds:	5	Gaps: 0																				
OR	<p>1 MetGlyGlnThrGlyLeuValSerGluValGlyProValCybTrpArgLeuValys 20</p>																					
DB	<p>93 ATGGGCCGACATCGGAGAATAATCGAAGAGGACAAGATTGTGGCCGAGACGCTATAAA 152</p>																					
OR	<p>21 SerGluTryMetcArgLeuArgGlnLeuLysArgPheArgGalaAlaSerGluValLysSer 40</p>																					

DB 153 TCAGAGTACATGCGACTGAGACAGCTCAAGAGCTTCAAGCAGCTGATGATGAAGTAAAGGT 212  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluValThrGluIleLeuAsnGlnGluTrp 60  
DB 213 ATGTTAGTTCACATCGTCAGAAAATTTGGAAACAGCAAAATCTTAAACCAAAAGTGG 272  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 273 AAACGCGAAGAGATACAGCTGTGTGACATCTTACCTCTGTAGCTCATGTGCGGAGACT 332  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 333 AGGAGGTGTTCGTGACCAAGTCACTTGAATTTTCCAAACAGATCATCCCATTAAGACT 392  
QY 101 LeuAsnAlaValAlaSerValProIleMetCysTrpSerTrpSerProLeuGlnGlnAspHe 120  
DB 393 CTGAATGACAGTGTCTTCAAGTACCCATAATGTATCTGTGTCTCCCTTACAGCAGAAATTT 452  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProCysMetGlyAspGluValLeuAsp 140  
DB 453 ATGTGTGAAGATGAACCTGTTTACATTAATCCGTATATGGAGATGAAGTATTAAC 512  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnIleLysAspGlyLysValHisIleGlyAsp 160  
DB 513 CAGAGTGTACTTTCATTTGAGAACTAATTAATAATTTATGATGAAAAGTACACGGGACT 572  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 573 AGAGATGTGGGTATTAATGATGAATTTTGTGGAGTGTGTCAATGCCCTTGTGTCA 632  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluGluGluGluGln 200  
DB 633 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 693 AAAGTCTGAGAGATCAACCGAGATATTAAGAAACCGCCCACTCGGAAATTTCTTCT 752  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 753 GATAAAATTTTGAAGCCATTTCTCGATGTTTCCAGATTAAGGCGACAGCAAGAACTA 812  
QY 241 LysGluLysIleLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGluCys 260  
DB 813 AAGGAAAATATTAAGAACTCAACGACAGCTCCAGGCGCACTTCTCTGAAGT 872  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 873 ACCCCCAATGATGAGACCAACCTTAATCTGTTCAAGAGACAAAGCTTACACTCC 932  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysIleAspCysPheLeuHisAspPheHis 300  
DB 933 TTTCAATACGCTTTCTGTAGGCGATGTTTAAATATGACTGCTTCTCATCTTTTCTCAT 992  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 993 GCACACCAACCAACGCTTATTAAGGGAAGAACACAGAAACAGCTTGAACAACAACTTGT 1052  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1053 GACACCAAGTGTACAGATTTGAGAGAGCAAGAGTTCGCTGCTCTCATCTCT 1112  
QY 341 GluArgIleLysThrProProLysArgProGlyArgArgArgGlyArgLeuProAsn 360  
DB 1113 GAGCGGATCAAGACCCCAACAAACGCGCAAGAGCCCAAGAAAGAGAGCGGCTTCCAT 1172  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1173 AACGATACAGGCGCCACAGACCCCACTTATATGCTGAGATCAAAAGATACAGACAGT 1232  
QY 381 AspArgGluAlaGlyThrGluThrGlyGluAsnAspLysGluGluGluGluGluGluGlu 400

DB 1233 GATAGGGAAGCAGGAGACTGAACCGGGGAGAGAAACATGATTAAGAAAGAAAGAGAG 1292  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGluThrProIleLysMet 420  
DB 1293 AAAGCGAACTTGGAGCTCTCTGAAAGCAAAATTCCTCGTGTCAAAACCAATTAAGATG 1352  
QY 421 LysProAsnIleGluProProGluLysValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1353 AAGCCAAATATTGAACCTCTCGAGATGTGAGAGTGTGTCTGAAGCTTCATATTT 1412  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1413 AGAGTCTCATTTGACACTTATGACAAATTTCTGTGCACTTGTGCTAGGTTATTTGGAGC 1472  
QY 461 LysThrCysArgGlnValTyrGluPheArgValIleLysGluSerSerIleIleAlaProAla 480  
DB 1473 AAAACATGTACACAGGTGTATGAATTTAAGATCAAAAGATTCATCATATGCTCAGCT 1532  
QY 481 ProAlaGluAspValAspThrProProArgLysLysValArgLysHisArgLeuThrAla 500  
DB 1533 CTGTGTAGATGTGACACTCTCTCCAGAGAAAAGAAAGAGAAACCGGTTGTGGCT 1592  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnIle 520  
DB 1593 GCACACTGCAAGAAAGATACAGCTGAAGAAAGAGACCGGCTCTTAACCATGTTTACACTAT 1652  
QY 521 GlnProCysAspHisAspArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1653 CAACCTGTGATCTTCAACCGGACCTTGTGACAGTGTGCTCCCTGTGTATATGACAA 1712  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1713 AATTTTGTGAAGTTTGTCAATGTATGTTCAAGTGTCAAAACCGTTTCCGGAGTCC 1772  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
DB 1773 CGCTGCAAGCAGACGTGCAACCAAGACAGTCCCGTCTACTGCGCGCTCGAGAGTGT 1832  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisAspAspSerLysAsnValSer 600  
DB 1833 GACCTGACCTCTCTCTTCACTTGTGAGCCGCTGACCACTTGGACAGTAAATAATGTGTC 1892  
QY 601 CysLysAsnCysSerIleGlnArgIleSerLysLysHisLeuLeuAlaProSerAsp 620  
DB 1893 TGCAGAACTGACATTTCAAGGAGGCTCAAAAGACATTTATGCTGCGCACATCTGAT 1952  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
DB 1953 GTGCAAGCTGGGGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 2012  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
DB 2013 TACTGTGACAGATTAATTTCAAGATGAAGCTCAAGAAAGGAAAGTATGATTA 2072  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
DB 2073 TACATGTGACAGCTTCTGTCACTTAACATATATTTGTGGAGATGACACCGGAG 2132  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
DB 2133 GGTAAACAATAATCGTTTGGCAATCATTCGTAAATCCAACTCTATGCAAAAGTT 2189

RESULT 7  
BD186247  
LOCUS BD186247 2558 bp DNA linear PAT 17-JUN-2003  
DEFINITION STATE activating gene.  
ACCESSION BD186247  
VERSION BD186247.1 GI:31878447  
KEYWORDS WO 02096943-A/116.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1 (bases 1 to 2558)  
AUTHORS Honda, G., Matsuda, A., Muramatsu, S. and Ishizawa, K.  
TITLE STR6 activating gene  
JOURNAL Patent: WO 02096943-A 116 05-DEC-2002;  
ASHBI KASBI CORP./GOICHI HONDA, AKIO MATSUDA, SHUJI MURAMATSU, KENYA ISHIZAWA  
COMMENT OS Homo sapiens (human)  
PN WO 02096943-A/116  
PD 05-DEC-2002  
PR 22-MAY-2002 WO 2002JP004949  
PR 25-MAY-2001 JP OIP 157043, 30-AUG-2001 JP OIP 260681 PR  
10-OCT-2001 JP OIP 313175  
PI GOICHI HONDA, AKIO MATSUDA, SHUJI MURAMATSU, KENYA ISHIZAWA PC  
C07K14/47, C07K17/00, C12N15/12, C12N5/10, C12P21/02, C12P21/08, PC  
C12Q1/02,  
PC A61P3/06, A61P3/10, A61P29/00, A61P31/00, A61P35/00, A61P37/00, PC  
GOIN33/15,  
PC GOIN33/50, GOIN33/53, GOIN33/566, G06F17/30, G06F17/60 CC STAT6  
activating gene  
FH Key Location/Qualifiers  
FT CDS Location/Qualifiers  
1..2558  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 2558  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.8% Indels: 0  
DB: 2 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x BD186247 (1-2558)

QY 1 MetGlyGlnThrGlyLysLeuSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
DB 58 ATGGGCCAGACTGGGAAAGAAATCTGAGAGGGAGCCAGTTTGTGGCGAAGCGTGTAAA 117  
QY 21 SerGlyrMetcArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 118 TCAAGGTCATGCGCATGAGCAGCTCAAGAGGTTCAAGCAGCGTGAAGTAAGTAAAGAT 177  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluValGlnGluIleLeuAsnGlnGluTrp 60  
DB 178 ATGTTTACGTTCCAAATCGTCAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 237  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 238 AAACAGCAAGGATACAGCGCTGTGCACATCTGACTTCTGTAGCTCATTGGCGGGACT 297  
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 298 AGGAGGTGTCGGTGACCAAGTGACTTGATTTTCCAAACAAAGTCATCCATTAAAGACT 357  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnGlnAsnIle 120  
DB 358 CTGAATGCAAGTTGCTTCAGTACCCCAATATGTAATTTCTTGCTCCCTTACAGCAGAAATTTT 417  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProLysMetGlyAspGluValLeuAsp 140  
DB 418 ATGGTGAAGATGAATCTGTTTACATTAACATTCCTTATATGGAGATGAAGTTTAAAT 477  
QY 141 GlnAspGlyThrPheIleGluGluLeuLysAsnLysArgGlyLysValHisGlyAsp 160  
DB 478 CAGAGTGGTACTTTTCAATTGAAGAACTAATTAATAATTAATGAGGAAAGTACACGGGAT 537  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180

DB 538 AGAAGATGCGGTTTATTAATGATGAATTTTGTGAGTGTGTAATGCCCTTGGTCAA 597  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLysGln 200  
DB 598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 201 LysAspLeuGluLysAspHisAspAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 658 AAAGATCTGAGAGATCAACCAATGATTAAGAAAGCGCCACCTCGGAAATTTCTTCT 717  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 718 GATTAATTTTGGAGGCCATTTCTCTCAATTTTCCAGTAAAGGCACAGCAAGAACTA 777  
QY 241 LysGluLysTyrLysGluLeuThrGlnGlnGlnLeuProGluValAlaLeuProGluCys 260  
DB 778 AAGGAAATATTAAGAACTCACCGAACACAGCTCCAGGCGCACTTCTCTCGAATGT 837  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 838 ACCCCCAACATGATGACCAAAATGCTAAATCTGTTCAGAGAGAGCAAAACCTTACCTC 897  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 898 TTTCATACGCTTTTCTGTAGCCATGTGTTTAATATACCTGCTTCTTCAATCTTTTCAT 957  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 958 GCACACCCCAACACTTATTAAGCGGAGAGACACAGAAACAGCTCTAGACAAACAACTTGT 1017  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGlyValAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1018 GGAACCAAGGTTTACCGCATTTGGAGAGGACAAAGAGATTTGCTGCTCTCACCGCT 1077  
QY 341 GluArgIleLysThrProProLysArgProGlyGlyValArgArgArgGlyArgLeuProAsn 360  
DB 1078 GAGCGGATTAAGACCCCAACAAACCTCCAGAGAGCCGCGAGAGAGAGACGGCTTCCCAT 1137  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleGluSerLysAspThrAspSer 380  
DB 1138 AACAGTAGAGGCCCAACACCCCACTTAATGTCTGAAATCAAGATACAGACAGT 1197  
QY 381 AspArgGluAlaGlyThrGlnThrGlyGlyGluAsnAsnAspLysGluGluGluLys 400  
DB 1198 GATGAGGAGCAGGAGCTGAACCGGGGAGAGAAATATTAAGAAAGAAAGAAAG 1257  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1258 AAAGATGAATCTGAGCTCTCTGAAAGCAAAATCTCGGTGTCAAAACCAATTAAGATG 1317  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1318 AAGCCAATATTTGAACCTCTCGAAGAAATGTGAGAGTGAGTGCTGAAGCTCAATGTT 1377  
QY 441 ArgValLeuIleGlyThrTrpTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1378 AGAGTCTCATGTGCACTTACTATGACAATTTCTGTCCATGTCTAGGTTAATGGAGCC 1437  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1438 AAACATGTAGACAGGCTGTATGAGTTAGACTCAAAAGAACTTAGCATCATAGCTCCAGCT 1497  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500  
DB 1498 CCGCTGAGATGATGATGATCTCTCCAGAGAAAGAAAGAGAAACACGGGTGTGGGCT 1557  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnThr 520  
DB 1558 GCACACTGCGAAAGAAATACAGCTGAAGAAAGACGGCTCTCTTAACATGTTTAAACATAT 1617  
QY 521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1618 CAACCTGTATCATCCACGGCAGCTTGTACAGTTCTGTCCCTTGTGTGATGACCA 1677

QY	561	AsnPhcCySGluLyPheCysGlnCySserSerGluCySGlnAsnArgPheProGluCyS	560
Db	1678	AATTTTGTGAAAAGATTTTGTCTCAAGTGAAGTGTCTCAAAACCGCTTTCCGGGAATGC	1737
QY	561	ArgCySLysAlaGlnCyAsnThrLySglnCySProCySTyrLeuAlaValArgGluCys	580
Db	1738	CGCTCCAAAGCACATGCAACACCAAGCAGTCCCTGTCTACTTGGCTGTCCGAGAGTGT	1797
QY	561	AspProAspLeuCyLeuThrCySGlyAlaAlaAspHisTrpAspSerLysAsnValSer	600
Db	1798	GACCTGTGACCTGTGTCTTACTTGTGGAAGCCGTGACCAATTGGGACAGTAAAAATGTGTCC	1857
QY	601	CysLyAsnCySserLleGlnArgGlySerLySlySHleuLeuLeuAlaProSerSerp	620
Db	1858	TGCAGAACTGCAAGTATTCAGGGGGGCTCCAAAAGCAATCTATGCTGCAACATCTTAC	1917
QY	621	ValAlaGlyTrpGlyLlePheLleLysAspProValGlnLysAsnGluPheLleSerGlu	640
Db	1918	GTCGACAGGCTGGGGGATTTTATTCAAAGATCCTGTGCGAAGAAAATGAATTATCTCAAA	1977
QY	661	TyrCySGlyGlnLleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys	660
Db	1978	TACTGTGAGAGATTAATTTCTCAAGATTAAGCTGCACAAGAGGGAAAGTGTATGATATA	2037
QY	661	TyrIleCySserPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys	680
Db	2038	TACATGTGCACGCTTCTGTCTTCAACTGTGAACAATGATTTTGTGTGTGGAGCAACCCGCAAG	2097
QY	661	GlyAsnLySlyLeuArgPheAlaAsnHisSerValAspProAsnCySTyrAlaLysVal	699
Db	2098	GGTAAACAAATTCGTTTGGCAATCATTCGGTAATCCAAATCGCTTATGCAAAAGTT	2154

RESULT 8				
LOCUS	AX821950	2576 bp	DNA	linear
DEFINITION	Sequence 78 from Patent WO03068961.			
ACCESSION	AX821950			
VERSION	AX821950.1	GI:39725171		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Homidae; Homo.			
REFERENCE	1			
AUTHORS	Andrews, P. A., Walsh, J. A. and Gokhale, P. A.			
TITLE	Method to modify differentiation of pluripotential stem cells			
JOURNAL	Patent: WO 03068961-A 78 21-AUG-2003;			

FEATURES	Location/Qualifiers
source	1..2576
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"

## ORIGIN

Alignment Scores:	
Pred. No.:	0
Score:	3802.00
Percent Similarity:	99.94
Best Local Similarity:	99.94
Query Match:	99.84
DB:	2
Length:	2576
Matches:	698
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

US-10-773-302-2 COPY 1 699 (1-699) X AX821950 (1-2576)

QY 1 MetGIyGInthrgIyLysLysSerGInLysGlyProValCystrPArgLysAsnValIys 20  
 Db 58 ATGGCCAGACTGGGAGAAATCTGAGAGAGGACCAAGTTTGTGGCGGAGACGGTAA 117  
 QY 21 SerGIyTyrMetAArgLeuArgGInLeuLysArgPheArgAlaAspGlnValLysSer 40

[illegible]

401 LysApGluThrSerSerSerSerSerGluAlaSerSerArgCysGlnThrProIleLeuMet 420  
1258 AAAGTGAACCTTCAGAGCTCCTCGAACCAATTCCTCGGTGCAACCAATTAAGATG 1317  
421 LysProAlaIleGluProProGluAlaGluTyrSerGlyAlaGluAlaSerMetPhe 440  
1318 AAGCAATATATTGAACCTCGAGAAATGTGGAGTGGAGTGGTGGTGAAGCTCAATGTTT 1377  
441 ArgValIleuIleGlyThrThrThrThrAspAsnMetPheAlaIleAlaArgLeuIleGlyThr 460  
1378 AGAGCCCATTTGGAGCTTACTATGACAAATTCGTGGCAATTCGTAGGTAAATGGAGCC 1437  
461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
1438 AAAACATATAGACAGGTATAGAGTTTAGAGTCAAAAGATCTAGATCATAGCTCAAGCT 1497  
481 ProAlaGluAspValAspThrProProAlaGlyLysLysLysLysLysLysLysLysLys 500  
1498 CCGGCTGAGATGTGATATCTCTCCAGAGAAAAGAGAGAAACACCGGTTGTGGGCT 1557  
501 AlaHisCysArgLysIleGluLeuLysLysAspGlySerSerSerAsnIleValTyrAsnTyr 520  
1558 GCACACTGCAGAAAGATACAGCTGAAAAGACGGCTCTTACCATGTTTACACTAT 1617  
521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
1618 CACCTGTGATCATCCACGGAGCTTGTGACATTTGTGCGCTTGTGTATGACACA 1677  
541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
1678 AATTTTGTGAAAATTTTGTCAATGTAGTTCAGAGTTCAAAACCGCTTCCGGAGAGC 1737  
561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLysAlaValArgLysCys 580  
1738 CGCTGCAAGCACAGTGCACCAAGAGTCCCGGTGTACTGCTGTCCGAGAGT 1797  
581 AspProAlaLeuCysLeuThrCysGlyValAlaAspHisStrPaspSerLysAsnValSer 600  
1798 GACCTGACCTGTCTTACTTGTGTGAGCGGTGACCATTTGGGACAGTAAATGTGTCC 1857  
601 CysLysAsnCysSerIleGlnArgLysLysLysLysLysLysLysLysLysLysLysLys 620  
1858 TGCAAGAACTGCAGATTCAGCGGGCTCCAAAAGACATCTATTCCTGGCACCACTGAC 1917  
621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
1918 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAAAATGAATTCATCTCAGAA 1977  
641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
1978 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAGAGGAGAAAGTGTATGATAA 2037  
661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
2038 TACATGTGACAGCTTCTGTCTCAACTGAAACATGATTTTGTGTGAGTCAACCCGCAAG 2097  
681 GlysLeuLysIleArgPheAlaAsnHisSerValAspProAsnCysTyrAlaLysVal 699  
2098 GGTAACAATAATCGTTTGCATAATCATTCGGTAAATCCAACTGCTATGCAAAAGTT 2154

RESULT 9  
HSZESTENH 2576 bp mRNA linear PRI 18-APR-2005  
LOCUS H.Sapiens mRNA for enhancer of zeste.  
ACCESSION X95653  
VERSION X95653.1 GI:1438063  
KEYWORDS zeste enhancer gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE 1 (bases 1 to 2576)  
AUTHORS Chen, H., Roessler, C. and Antonarakis, S. E.  
TITLE Cloning of a human homolog of the Drosophila enhancer of zeste gene (EZH2) that maps to chromosome 21q22.2  
JOURNAL Genomics 38 (1), 30-37 (1996)  
PUBMED 8954776  
REFERENCE 2 (bases 1 to 2576)  
AUTHORS Antonarakis, S. E.  
TITLE Direct Submission  
JOURNAL Submitted (13-FEB-1996) S. E. Antonarakis, Division of Medical Genetics, University of Geneva School of Medicine, Rue Michel-Servet 1, CH-1211 Geneva, SWITZERLAND  
COMMENT Related protein P42124.  
FEATURES  
source  
location/Qualifiers  
1..2576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="q22.2"  
/tissue\_type="brain"  
/dev\_stage="fetal"  
1..2576  
/gene="EZH2"  
58..2298  
/gene="EZH2"  
/codon\_start=1  
/product="enhancer of zeste"  
/protein\_id="CA64955.1"  
/db\_xref="GI:1438064"  
/db\_xref="GOA:Q15910"  
/db\_xref="UniProtKB/Swiss-Prot:Q15910"  
/translation="MGQTGRKSEKGPVCKRRKRVSEYRRLQLKRRFRADVKMFSS  
NRQKILRTEILNQEMKRIQPVHILTVSLSLGTRECSYTSDDLPPTQVILPKTLN  
AVASVPIYMSWSPLQONFVEDTVLNI PYMGDEVLDDGTFIEELIKYDGVHGD  
REGFINDEIFVELVNALGGYNDDDDDDDGDDPERREKODLEDHDDKESRPRK  
PSDILALISMPDKGTABEIKKEYELRBOCLRALPRCPNIDGPAKSVOREQ  
SLSHFHTLPCRRKTYDCLFHPHATNTYTKRKTETRLADLNKPGPCYCHLEAKRF  
AAALTAERIKTPPRPGRRGRRLPNSSSRSTLTIVLSKDDTSREAGTERGEN  
NDKEERKQETSSSSBANSRCQPIPMKQNIIEPENVWSGASMEFVLIIGYIYN  
FCALRLIGTTCGOVYEFVYKESIAPAPAEVDVPPRKKKKRLMAHCKRIOL  
KDGSSNMHVNYOPCDHPRODCSSCPVIAQNFCEKSCQSSCONRFPCKCKAOC  
NTKQPCYLAVEREDPDLCTCCGAADMDKNSCKRCSIQRSKKHLLALPSVAVG  
GIFLKDVPQKNEFISEYCGELTISDEADRKRYDKNCTMCSFLPLNLDPAVDRKEN  
KIRPANSHVNPNCYAKVMVNGDHRIQIFARALQTSBELFVDYRYSQADALKYGLIS  
REMEIP"  
2558  
/gene="EZH2"

ORIGIN  
polyA\_site  
alignment Scores:  
Pred. No.: 0 Length: 2576  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.8% Indels: 0  
Gaps: 0

US-10-773-302-2\_copy\_1\_699 (1-699) x HSZESTENH (1-2576)

401 MetGlyGlnThrArgLysLysSerGluLysGlyProValCysTyrPargLysArgValLys 20  
58 ATGGCCGACAGCTGGGAAATAATCGAGAAAGGACCAAGTTTGTGGCGAAGCGGTAA 117  
21 SerGluTyrMetArgLysArgGlnLeuLysArgPheArgValAlaAspGluValLysSer 40  
118 TCAAGATCATGCGACTAGACAGCTCAAGAGGTTTCAACAGCTGATGAAGTAAAGCT 177  
41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnGluTyr 60  
178 ATGTTTACTTCCATTCGTCAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAAATGG 237  
61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80

Db 238 AAGACGGAAGATACGCTGTGACATCTTCTGTGAGCTCATTTGGCGGAGCT 297  
 Qy ArgValLeuIleGlyThrTyrTrpAspPheCysAlaIleAlaArgLeuIleGlyThr 100  
 Db 298 AGGAGTGTTCGGTGACCAAGTGCATGATTTTCCAACACAAAGTCATCCATTAAGACT 357  
 Qy LeuAspAlaValAlaSerValProIleMetTyrSerTyrSerProIleuIngnanphe 120  
 Db 358 CTGAATGCAGTGTCTCAGTACCACTTAATGATATTTGTGCTCTCCCTACGCGAATTTT 417  
 Qy 121 MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleAsp 140  
 Db 418 ATGTGTGAAGATCAAACTGTTTACATTAATCTCTTAATATGGAGATGAAGTTTAAAT 477  
 Qy 141 GluAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
 Db 478 CAGGATGTATCTTTCATTTGAAAGAACTAATTAATAAATTTATGATGGAAAGTACACGGGAT 537  
 Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
 Db 538 AAGAAATGTGGTGTATTAATGATGAATTTTGTGAGTGTGTGAATGCCCTTGTGTCAA 597  
 Qy 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
 Db 598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 Qy 201 LysAspLeuGluAspHisArgAspAspGlyGluSerArgProArgGlyPheProSer 220  
 Db 658 AAAGATCTGAGGATCACCGAGATGATTAAGAAAGCCGCCACCTCGAAATTTCTTCT 717  
 Qy 221 AspValIlePheGluAlaIleSerSerMetPheProAspGlyGlyThrAlaGluGluLeu 240  
 Db 718 GATTAATTTTGGAGGCCATTTCTCAATGTTTCCAGATAGGGCCACAGAGAACTTA 777  
 Qy 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyValAlaLeuProGluLys 260  
 Db 778 AAGAAATATTAAGAACTCACCGAACGACCTCCAGCGCACTTCTCTTGATAT 837  
 Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
 Db 838 ACCCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
 Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
 Db 898 TTTCATACCTTTTCGTAGCGATGTTTAAATGATGATGATGATGATGATGATGATGATGAT 957  
 Qy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 Db 958 GCAACACCCACACTTATTAAGCGAAGAACACAGAACAGCTCTAGACAAACAACTTGT 1017  
 Qy 321 GlyProGlnCysTyrGlnHisLeuGluGluLysAlaLysGluPheAlaAlaIleLeuThrAla 340  
 Db 1018 GGACCAACAGTGTATACAGCATTTTGGAGGGAAGGAATTTGCTGCTGCTCTCACCGCT 1077  
 Qy 341 GluArgGlyLeuLysThrProProLysArgProGlyGlyLysArgArgGlyValArgLeuProAsn 360  
 Db 1078 GAGCGGATTAAGACCCCAACAAACGTCACAGAGCGCGACAGAGAGAGCGGCTTCCCAT 1137  
 Qy 361 AsnSerSerArgProSerThrProThrIleAsnValIleGluSerLysAspThrAspSer 380  
 Db 1138 AACAGTAGAGGCCCAACAGCCCAACATTAATGTGCTGGAATCAAAAGATTCACAGAGT 1197  
 Qy 381 AspArgGluAlaGlyThrGluThrGlyGluAsnAsnAspLysGluGluGluLys 400  
 Db 1198 GATTAGGGAAGCAAGGAGCTGAACCGGGGAGAGAACATGATTAAGAGAGAGAGAGAGAG 1257  
 Qy 401 LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 Db 1258 AAAGATGAACCTTCAGGCTCCCTCTAAGCAAAATTCCTCGGTGCAAAACCAATTAAGATG 1317  
 Qy 421 LysProAsnIleGluProProGluAsnValGluTyrPserGlyValGluAlaSerPhe 440

Db 1318 AAGCAAAATATTAAGCTCTGAGAAATGAGAGTGGAGTGGTGTGAGCTCAAGCTCATTTT 1377  
 Qy 441 ArgValLeuIleGlyThrTyrTrpAspPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1378 AAGATCTTCATGAGCACTTACTATGACATATTTCTGTGCAATGATGATTAATTTGGAGAC 1437  
 Qy 461 LysThrCysArgGluValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 Db 1438 AAAACATGTAGACAGGTGTATGATGATTAAGTCAAAAGAAATCTAGCATCACTCAGCT 1497  
 Qy 481 ProAlaGluAspValAspThrProProArgLysLysArgLysHisArgLeuThrAla 500  
 Db 1498 CCGGTGAGAGATGTGATCTCTCTCAAGAGAAAGAGAGAGAAACCGGTGTGGGCT 1557  
 Qy 501 AlaHisCysArgLysIleGluLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 Db 1558 GCACACTGACGAAAGATACAGCTGAAAAGAGCGCTCTCTACACATGTTTAAACATAT 1617  
 Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 1618 CAACCTGTATCATCACGCGAGCTTGTGACAGTGTGTGCTGTGTGTATGACACAA 1677  
 Qy 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 Db 1678 AATTTTGTGAAGATTTTGTCAATGTATGATGATGATGATGATGATGATGATGATGATGAT 1737  
 Qy 561 ArgCysLysValGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys 580  
 Db 1738 CGCTGCAAGCACAGTGCACACCAAGAGTCCCGCTGCTACCTGCTGTCCGAGAGTGT 1797  
 Qy 581 AspProAspLeuCysLeuThrCysGlyValAlaAlaAspHisTrpAspSerLysAsnValSer 600  
 Db 1798 GACCTGACCTTGTGCTTATCTGTGAGCCGCTGACATGTGGACAGATTAATATGTGCTC 1857  
 Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuLeuAlaProSerAsp 620  
 Db 1858 TGCAGAACTGCAAGTATTCAGGGGGCTCCAAAAGCATCTATTGCTGGACCATCTGAC 1917  
 Qy 621 ValAlaGlyTyrPheLysPheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1918 GTGCGAGCTGGGGGATTTTATCAAGATCTGTGCAAAAATGAATTCATCTCAGAA 1977  
 Qy 641 TyrCysGlyGluIleIleSerGluAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 Db 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAGAGTGTATATATAA 2037  
 Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
 Db 2038 TACATGTGACGCTTCTCTTCACTTGAAACATGATTTTGTGTGATGACCAACCGGAG 2097  
 Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 Db 2098 GGTAACAAAATTCGTTTGAATCAATCATTCGGTAATTCAAACTGCTATGACAAAAGTT 2154  
 Db 2098 GGTAACAAAATTCGTTTGAATCAATCATTCGGTAATTCAAACTGCTATGACAAAAGTT 2154  
 RESULT 10  
 BC010858 2655 bp mRNA linear PRI 02-JUL-2005  
 LOCUS BC010858  
 DEFINITION Homo sapiens enhancer of zeste homolog 2 (Prosoephila), transcript  
 variant 1, mRNA (cDNA clone MGC:9169 IMAGR:3901250), complete cds.  
 ACCESSION BC010858  
 VERSION BC010858.2 GI:34194096  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominoidea; Homo.  
 REFERENCES  
 1 (bases 1 to 2655)  
 Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Alech, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,





Db	903	ACCCCAACATGATGAGACCAAAATCTAAATCTGTCAGAGAGCAAAAGCTTACATCC	962
OY	281	PheIsthrLleupheCyatgthgCyabheIyTyaaPcybheLuehi-----	297
Db	963	TTTCATACGCTTTCTGTATGGCGAATGTTTAAATATATGATCTGCTCATCTGTAAATGTC	10222
OY	288	-----ProPhehiAAlaThrProasnthTytyIybaIybaantThcIuThraIaleu	315
Db	1023	AATTAATCTTTTCATGACAAACCCCAACCTTATTAAGCGAAGAAACAGAAACAGCTCTTA	10822
OY	316	AspaenIySProCYaGIyProGInCystrYtGInhiSleuGInGyAlaIyGInupheIa	335
Db	1083	CACAAACAACCTTGTGGACACACAGTGTATCCAGCATTTGGAGGAGCAAAAGATTTCT	11422
OY	336	AlaAlaIeuthraIaGIuagTgIleIystrThrProProlYbaIyProGIyGIyATgthgth	355
Db	1143	GCTGCTTCACCGCTGAGCGGATTAAGACCCCAACAAACGTCCAGAGGCGCGAGAGA	12022
OY	356	GIyATgIeupProabnsnSerSerATgProserThrProthrlIeasnaIleuGInser	375
Db	1203	GGACGGCTCTCCAAATMAACATGACGAGCGCCACAGACCCCAACATTAATGTCTGGAAACA	12622
OY	376	LyasapThraapSerAspaArgGluabIythrGIuThrGIyGIuIaasnaAspIy	395
Db	1283	AAAGATACAGACAGATATGAGGAGACAGAGACTGAACGGGGGGAGAAACAATGATATA	13222
OY	396	GIuGIuGIuGIuIyIyIyAspGIuThrSerSerSerSerGIuIaIaasnaSerATGyGIuIn	415
Db	1323	GAAAGAGAGAGAGAAAGATGAATCTTGAGCTCTCTGAAGCAAAATTCCTGGGTGCA	13822
OY	416	ThrProIleIyMetIyAspProasnlIeGIuProProlYbaasnaIuGIuThrSerGIyAla	435
Db	1383	ACACCAATMAAGATGAACCAAAATATGAACTCTCTGAAATGTGGAATGAGATGAGTGTCT	14422
OY	436	GIuIaIasermecPheATgValIeulIeGIyThrTytyTaaasnaPheCyAlaIleIa	455
Db	1443	GAAAGCTCAATGTTTAAGATGCTCTCATTTGACCTTATGACAAATTTCTGAGCAATGCT	15022
OY	456	ArgIeulIeGIyThrIyThrThrCyAspGIuIaIyTgIuPheATgValIyGIuSerSer	475
Db	1503	AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAAGTCAAAAGATCTAGC	15622
OY	476	IleIleIaIaProlaIaProlaGIuAspValaAspThrProProlYbaIyIybaIygaIy	495
Db	1563	ATCATAGCTCAGCTCCCGCTGAGGATGTGATACTCTCCAAAGGAAAGAAAGAGGAAA	16222
OY	496	HisATgIeuthraIaAlaIhiSicCyATgIyGIuIeGIuIeulIyIyIybaGIySerSerAsn	515
Db	1623	CACCGGTGTGTGGCTGACACTGCAAGAAATACAGCTGAAAGAAAGAGAGGCTCTCTAAC	16822
OY	516	HisValTyTaaanTyGIuIaProCYaAspIaAspIaProlATgGIuIaProCYaAspSerSerCyAspPro	535
Db	1683	CATGTTTCAACTATCAACCTCTGTATATCAAGGACGCTTGTGAACATGTCGTCCT	17422
OY	536	CybaValIlealagIaasnaPheCyGIuIyAspheCyGIuIySerSerGIuIyGIuIaAsn	555
Db	1743	TGTGTGATAGACAAATATTTTGTGAAAGATTTTGTCAATGATGATTCAGATGTCAAAAC	18022
OY	556	ArgPheProlYbaIyCybaIyGIuIaIaGIuIyIybaasnaThrlYGIuIyProCYaTyIeIu	575
Db	1803	CGCTTTCCGGGATGCGCTGCAAAAGCAACAGTGCAACCAAGACAGTGCCTCGTGTAACCTG	18622
OY	576	AlaValATgGIuIyCybaAspProAspIeulCybaIeuthrCyGIuIaIaIaAspIaIaAspIa	595
Db	1863	GCTGTCCAGAGTGTGACCTCTGACTCTGTCTTACTTGTGAAGCGCTGACATTTGGGAC	19222
OY	596	SerIybaasnaIaSerCyAlaIybaasnaCySerIleGIaATgGIySerIyIybaIeIuIeIu	615
Db	1923	AGTAAATATGTCTCTGCAAGAACTGCAAGTATTCAGCGGGGCTCCAAATAACATCTATTTG	19822
OY	616	IeulIaIaProlSerAspValAlaGIyTThrGIyIlePheIleIybaAspProValGIuIyAsn	635
Db	1983	CTGGGACATCTGACGTGACGCTGCGGGGATTTTATCAAAAGATCTCTGTGCAAGAAAT	20422

QY		636	GluPhelIleSerGlnUrrCysGlyValIleLeuSerGlnPheGluAlaAspArgArgGly	655
Db		2043	GAATTCACTCAGAAATACGTGGAGAAATTAATTTCTCAAGATGAAGCTGCACAGAAGCGG	2102
QY		656	LysValTyrAspLysTyrMetCysSerPheLeuPheAseLeuAseAnaAspPheValVal	675
Db		2103	AAAGGTATGATAAATACATGTGACGCTTCGTCTCAACTTGAAACATGATTTTGTGTGTG	2162
QY		676	AspAlaThrArgLysGlyAseMylLeaArgPheAlaAseHisSerValAseProAseCys	695
Db		2163	GATCAACACCSCCAAGGGTAAACAATAATCGTTTGCCAATCATTCGTAAATCCAAATCGC	2222
QY		696	TyrAlaLysVal 699	
Db		2223	TATGCCAAAAGTT 2234	
RESULT 11				
LOCUS	DD208658	2512 bp	DNA	linear PAT 19-JUN-2006
DEFINITION	Expression Profile of Prostate Cancer.			
ACCESSION	DD208658			
VERSION	DD208658.1 GI:85649860			
KEYWORDS	JP 200518522-A/95.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Bukacinski, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominales; Homo. 1 (bases 1 to 2512) Rabin,M.A., Chinaiyan,A.M. and Surikumaru,A. Expression Profile of Prostate Cancer Patent: JP 200518522-A 95-23-JUN-2005. THE REGENTS OF THE UNIVERSITY OF MICHIGAN			
AUTHORS	OS Homo sapiens			
JOURNAL	PN JP 200518522-A/95			
COMMENT	PD 23-JUN-2005			
	PF 02-AUG-2002 JP 2003517245			
	PR 02-AUG-2001 US 60/330581,15-NOV-2001 US 60/334468, PR			
	01-AUG-2002 US 10/210120			
FEATURES	PI mark a rabin, arun m chinaiyan, arun surikumaru CC			
	PH Key Location/Qualifiers.			
SOURCE	Location/Qualifiers			
	1..2512			
	/organism="Homo sapiens"			
	/mol_type="unasigned DNA"			
	/db_xref="taxon:9606"			
ORIGIN				
Alignment Scores:				
Pred. No.:	0	Length:	2512	
Score:	3781.00	Matches:	698	
Percent Similarity:	99.9%	Conservative:	0	
Best Local Similarity:	99.9%	Mismatches:	1	
Query Match:	99.3%	Indels:	1	
DB:	2	Gaps:	0	
US-10-773-302-2_COPY_1_699 (1-699) x DD208658 (1-2512)				
QY	1	MetGlyGlnThrGlyLysLysSerGlnUrrGlyProValCysTrpArgLysArgValLys	20	
Db	10	ATGGGCCAGACTGGGAGGAATCTGAGAAAGGACCAAGTTGTCGCGAAGCGGTAAAA	69	
QY	21	SerGlnUrrMetArgLysArgGlnLeuLysArgPheArgArgAlaAspGlyValLysSer	40	
Db	70	TCAAGTAGCACATGACACAGACAGCTCAAGAGGTCTACAGACCTGATGAGTAAGTAAAGCT-	128	
QY	41	MetPheSerSerAsnArgGlnLysIleLeuGlnUrrGlnUrrLeuAseGlnGlnUrr	60	
Db	129	ATGTTTAGTTCCATGCTCAGAAAATTTTGGAAAAGAACGAAATCTTAAACCAAGATGC	188	
QY	61	LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr	80	

Db 189 AACACGCGAAGATACAGCCTGTGCACATCTGTGACCTTGTGAGCTCATTTGGCGGAGCT 248  
 Qy ArgGluCySerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
 Db 249 AGGAGAGCTTGGGTGACCAAGTCACTTGATTTTCCAAACACAGATCATCCATTAAGACT 308  
 Qy LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAspPhe 120  
 Db 309 CTGAATGCACTGGCTTCAGTACCCCAATATGTAATCTTGGTCTCCCTTCACACAGAAATTTT 368  
 Qy MetValGlnAspGluThrValLeuHisAsnIleProYrMetGlnAspGlnValLeuAsp 140  
 Db 369 ATGGGGAAGATGAACTGTTTACATACATCTCTTATATGGAGATGAAGTTTACAT 428  
 Qy GlnAspGlyThrPheIleGluGluLeuIleLysAsnYrAspGlyLysValHisGlyAsp 160  
 Db 429 CAGATGCTACTTTTCATTGAAGAACTAATAAAAAATTATGATGGGAAAGTACACGGGAT 488  
 Qy ArgGluCySerGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
 Db 489 AGAGATGTGGGTATATGAATGAATGAAATTTTGTGGAGTTGGTGAATGCCCTTGTGCA 548  
 Qy TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLysGln 200  
 Db 549 TATATATGATGATGACATGATGATGATGACATCTCTGAAGAAAGAAAGAAAGACG 608  
 Qy LysAspLeuGlnAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
 Db 609 AAAAGATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAAATTTCTTCT 668  
 Qy AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
 Db 669 GATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGACGACGAGAAAGACTA 728  
 Qy LysGluLysYrTrpLysGluLeuThrGluGlnGluLeuProGluValAlaLeuProGluCys 260  
 Db 729 AAGGAAATATTAAGAACTCACCGAACGACGCTCCAGGGCGCATCTCTCTGAAATGT 788  
 Qy ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
 Db 789 ACCCCCAACATGATGAGACCAATGCTAATCTGTTCAGAGAGCAAGCAAGCTTACACTCC 848  
 Qy PheHisThrLeuPheCysArgArgCysPheLysYrAspCysPheLeuHisProPheHis 300  
 Db 849 TTTCAATGCGCTTTCTGTAGCGCATGTTTAAATATGACTGCTCTCTCATCTCTTTTCA 908  
 Qy AlaThrProAsnThrTrpLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 Db 909 GCACACCCCAACATTAATAGCGGAAAGACACAGAAACGCTTAGACACCAACCTTGT 968  
 Qy GlnProGlnCysYrGlnHisLeuGluGluValAlaLysGluPheAlaAlaAlaLeuThrAla 340  
 Db 969 GGACCAACAGTGTTCACAGCATTTGGAGGGAGCAAGGAGTTTGCCTGCTCTCACCGCT 1028  
 Qy GlnArgGlnIleLysThrProProLysArgProGluGlyArgArgArgGlyValArgLeuProAsn 360  
 Db 1029 GAGCGGATTAAGACCCCAACAAACGTCACGAGAGCCCGACAGAAAGGACGCGTCTCCAT 1088  
 Qy AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 Db 1089 AACAGTACAGGCGCCACGCCACCCACCATTAATGTGCTGGAATCAAGAGATACACAGCT 1148  
 Qy AspArgGlnAlaGlyThrGlnThrGlnGlyLysAsnAsnAspLysGluGluGluLys 400  
 Db 1149 GATAGGGAAGACAGGACGTAAGAACGGGGGAGGAACATGATTAAGAAAGAAAGAAAGAG 1208  
 Qy LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 Db 1209 AAAAGATGAACCTTCAGAGCTCCCTCAAGACAAATTCCTCGGTGTCAAAACCATATAAGATG 1268  
 Qy LysProAsnIleGluProProGluLysValGluTrpSerGlyValGluAlaSerMetPhe 440  
 Db 1269 AAGCCAAATATTAATCTCTCAAGAAATGTGAGATGTGTGTAAGCTTCATATGTTT 1328

Qy 441 ArgValLeuIleGlyThrYrTrpAsnAspPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1329 AGAGTCTCTCATTTGGCACCTTACTATGACAAATTTCTGTGCCATTTGTAGTAAATTTGGAC 1388  
 Qy LysThrCysArgGlnValYrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 Db 1389 AAAACATGTAGACAGGTGATGAGTTAGAGTCAAGAAATCTAGATCATAGCTCCAGCT 1448  
 Qy ProAlaGluAspValAspThrProProArgLysLysLysValArgLysHisArgLeuTrpAla 500  
 Db 1449 CCGGCTAGAGATGTGATTACTCTCCAAAGGAAAGAAAGAAACACCGGTTGTGGCT 1508  
 Qy AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValYrAsnYr 520  
 Db 1509 GCACATGCAAGAAAGATACAGCTGTAAGAAAGACGCTCTCTTAACCATGTTTACACTAT 1568  
 Qy GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 521 CAACTGTGATCATCCACGGACGCTTGTGACAGTTCGCTTGTGATGACACMA 1628  
 Qy AsnAspPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 Db 1569 CAACCTGTGATCATCCACGGACGCTTGTGACAGTTCGCTTGTGATGACACMA 1628  
 Qy 541 AsnAspPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 Db 1629 AATTTTGTGAAAGTTTGTCAATGTATGATTCAGAGTGCMAACCGCTTCCGGATGC 1688  
 Qy ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrTrpLeuAlaValArgLysCys 580  
 Db 561 CGCTGCAAGCACAGTGCACCAACCAACAGTCCCGGTCTTACCTGATCCGAGAGTGT 1748  
 Qy AspProAspLeuCysLeuThrCysGlyValAlaAspHisTrpAspSerLysAsnValSer 600  
 Db 1749 GACCTGATCTCTCTTACTTGTGAGCCGCTGACATTTGGACATTAATAATGTCTCC 1808  
 Qy CysLysAsnCysSerIleGlnArgLysTrpLysHisLeuLeuLeuAlaProSerAsp 620  
 Db 1809 TGCAAGAACGCAATTAATGACGGGGGCTCCAAAAGATCATATGTCGACCATCTGAC 1868  
 Qy ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1869 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAGAAATGAATTCATCTCGAA 1928  
 Qy TyrCysGlyGluIleIleSerGlnAspGlnAlaAspArgArgGlyLysValYrAspLys 660  
 Db 1929 TACTGTGAGATTAATTTCTCAAGATGAACTCAAGAAAGGAAAGTATGATTA 1988  
 Qy TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
 Db 1989 TACATGTGACAGCTTCTGTTCAACTTGAAACATGATTTTGTGGATGCAACCCGCAAG 2048  
 Qy GlnLysAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysYrTrpAlaLysVal 699  
 Db 2049 GGTAAACAAATTCGTTTGGCAAAATCAATCGGTAAATCCAACTGTGCAAAAGTT 2105

RESULT 12  
 BC003772 2574 bp mRNA linear ROD 29-JUN-2004  
 LOCUS Mue musculus enhancer of zeste homolog 2 (Drosophila), mRNA (cDNA clone MGC:5969 IMAGS:3492110), complete cds.  
 ACCESSION BC003772  
 VERSION BC003772.1 GI:13277755  
 KEYWORDS MGC.  
 SOURCE Mue musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Mus.  
 1 (bases 1 to 2574)  
 Strauberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Sherman, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Helel, F., Diatchenko, L., Marsina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T. B., Brownstein, M. J., Usdin, T. B., Toshiyuki, S.,  
Carnini, P., Prange, C., Raha, S., Loquellano, N. A., Peters, G. J.,  
Abramson, R. D., Mahony, S. J., Bosak, S. A., McEwan, P. J.,  
McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S.,  
Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W.,  
Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A.,  
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y.,  
Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D.,  
Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M.,  
Butterfield, Y. S., Krzywinski, M. I., Skalska, V., Smal, D. E.,  
Schnerch, A., Schein, J. E., Jones, S. J., and Marra, M. A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2574)  
Strausberg, R.  
Direct Submission  
Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaphs-remail.nih.gov](mailto:cgaphs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdpaxil.stanford.edu](mailto:mcdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRMA Plate: 8 Row: P Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6679720.  
Location/Qualifiers

## FEATURES

source

1..2574  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech 11"  
/db\_xref="taxon:10090"  
/clone="MGC:5969 IMAGE:3492110"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor  
arose spontaneously from a senescent normal mammary  
clone. 11b="NCI CGAP Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-Sport6"  
1..2574  
/gene="Ezh2"  
/note="synonym: Ezh-1"  
/db\_xref="GeneID:14056"  
/db\_xref="MGI:107940"  
54..2294  
/gene="Ezh2"  
/codon\_start=1  
/product="enhancer of zeste homolog 2"  
/protein\_id="AAH03772.1"  
/db\_xref="GI:13277756"  
/db\_xref="GeneID:14056"  
/db\_xref="MGI:107940"  
/translation="MGCTGKSKBPVCKRKYSEYRLNQLRFRADRYKTPSS  
NRKILERTLNQKRIQPVHMTSVSLRKRECVTSIDLPAYVPLKTLN  
AVASVPIVSWPLQONFVNEDEVLHNI PYMGDEVLDQGTIELIKNDGVKVG  
RBCGFINDIEIPELVNALGVYDDDDDDGDDPDERBKQKLEBNRDKTCPRKF  
PADKIFPAISSWFPDKGTAELEKRYKELTEQQLFGALPPECTPNI DGPNAVSREQ  
SLHSFHTLCRCRCFYDCFLHPFHTPMTYRKNTETALDNKPGCPGCGYHLEAKER

## gene

CDS

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2574  
Score: 3756.00 Matches: 687  
Percent Similarity: 99.38 Conservative: 7  
Best Local Similarity: 98.38 Mismatches: 5  
Query Match: 98.64 Indels: 0  
Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x BC003772 (1-2574)  
QY 1 MetGlyGlnThrGlyLysSerGlyProValCysTrpArgLysValYs 20  
DB 54 ATGGGCCAGACTGGGAGAGAAATCTGAGAGGACCGGTTTGTGGCGGAGGTAAAA 113  
QY 21 SerGluYrMetArgLysArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 114 TCAAGATCAATGAGACTAGACAGCTCAAGAGTTCAAGAGCTGATGAAGTAAAGACT 173  
QY 41 MetPheSerSerAsnArgGlnLysLeuGluArgThrGluLeuAsnGlnLysTrp 60  
DB 174 ATGTTAATTCCTCAATCGTCAGAAATTTTGGAAAGAACGAAACCTTAACCAAGATGG 233  
QY 61 LysGlnArgGlnLeuProValHisLeuThrSerValSerSerLeuArgLysTrp 80  
DB 234 AAGCAGCGAGAGATACAGCTGTGTCATCACTTCTGTGAGCTCATTTGGCGGAGCT 293  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuYrThr 100  
DB 294 AGGAGATTCATGATCAGACAGTGAATTTTCCAGCAAGATATCCGTTAAAGACC 353  
QY 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB 354 CTGAATGAGTGGCTGCTGCTGCTATATGATCTTGTGCTGCTTCAACAGAAATTTT 413  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProYrMetGlyAspGluValLeuAsp 140  
DB 414 ATGGTGAAGACGAAACGTCTTACATTAACATTTCTTACATGCGGAGATGAAGTTCTGAT 473  
QY 141 GlnAspGlyThrPheIleGluGlnLeuIleLysAsnYrAspGlyLysValHisGlyLysP 160  
DB 474 CAGGACGGCAGCTTCATTAAGAACTAAATTAATTAATTAATTAATTAATTAATTAATTA 533  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluGln 180  
DB 534 AGAGATGTGATTTATTAATGAATGAATTTTGTGAGTTGTGTAAGTCACTTCTGCTCA 593  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluGln 200  
DB 594 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 653  
QY 201 LysAspLeuGluIleAspAspAspAspAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 654 AAAGATTCAGAGATTAATGAGATTAATGAGATTAATGAGATTAATGAGATTAATGAGATTA 713  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 714 GATAAATAATTTTGAAGCAATTTCTCAATGTTTCAAGTAAAGGAGACCGAGAAAGAACTG 773  
QY 241 LysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 260  
DB 774 AAGGAAAAAATTAAGAACTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 833  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280





Db 1387 AGAGTCTCTCATGTGACTTACTAGCAATTAATTTGTGTCATGCTAGCTTAATTTGGAACC 1446  
Qy 461 LysThrCysArgGlnVal1TyrgLuphArgVal1LysGluSerSerIleIleAlaProAla 480  
Db 1447 AAAACATGTGACAGGTGTATGAGTTTGAAGTCAAGAGTCAATTAATTAATGACCTGTT 1506  
Qy 481 ProAlaGluSerVal1AspThrProProArgLysLysArgLysHisArgLeuTyrAla 500  
Db 1507 CCCACTGAGGATGTAGACACTCTCCAAAGAAAGAAAGAAACATCGGTGTGGGCT 1566  
Qy 501 AlaHisCysArgLysIleGlnLeuLysArgLysSerSerAsnHisVal1TyraAsnTyr 520  
Db 1567 GCACACAGCGAAGAAAGATCAACTGAAAGACCGCTCTCTTAACATGTTTACACTAT 1626  
Qy 521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysVal1IleAlaGln 540  
Db 1627 CAACCTGTGACATTCACCGGACGCTTGTACAGTTCTGCTCTTGTGTGATAGCACAA 1686  
Qy 541 AsnAspCysGluLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
Db 1687 AATTTTGTGAAAAGTTTGTGCATGTAGTTCAAGGTGCAAAACCGCTTCTCGAGATGT 1746  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrIleAlaVal1ArgLysCys 580  
Db 1747 CGGTGCAAGACACATGCAACCAACAGTGTCAATGTCACTGCTGCTCGAGAGTGT 1806  
Qy 581 AspProAspLeuCysLeuThrCysGlyVal1AlaAspHisTyrAspSerLysAspVal1Ser 600  
Db 1807 GACCTTGAACCTCTCTCTCACTGTGTGAGCTGTCAACATTTGGACAGTAAAAATGTATCC 1866  
Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuLeuAlaProSerAsp 620  
Db 1867 TGTAGAACTGTAGACATTCAGCGGGCTCTAAAAAGCACTTACGTGCGACCGCTCATGT 1926  
Qy 621 Val1AlaGlyTyrPglYllePheIleLysAspProVal1GlnLysAsnGlnPheIleSerGlu 640  
Db 1927 GTGGCAGGCTGGGGCATCTTTATCAAAAGATCCTGTACAAAAGAAATGAAATTCATCTCAGAA 1986  
Qy 641 TyrCysGlyGluIleIleSerGlnAspGlnAlaAspArgGlyLysVal1TyrAspLys 660  
Db 1987 TACTGTGGGAGATTTATTTCTCAGATGAAACGACAGAAAGAAAGTTATAGACAA 2046  
Qy 661 TyrMetCysSerPheLeuPheAsnLysAsnAspPheVal1ValAspAlaThrArgLys 680  
Db 2047 TACATGTGCACTTCTGTCAACTGAAACATGATTTTGTGAGTGAAGCAACCGGAAAG 2106  
Qy 681 GlysLeuLysIleArgPheAlaAsnHisSerVal1AsnProAsnCysTyrAlaLysVal 699  
Db 2107 GGCAACAAATTCGTTTGTCTAATCATTCAGTAAATCCAACTGCTATGCAAAAGTT 2163

RESULT 14  
AK220174 2600 bp mRNA linear ROD 01-MAR-2005  
LOCUS AK220174  
DEFINITION Mus musculus mRNA for mKIAA4065 protein.  
ACCESSION AK220174  
VERSION AK220174.1 GI:60360229  
KEYWORDS FLI CDNA.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Okazaki, N., Kikuno, R.F., Ohara, R., Inamoto, S., Koseki, H.,  
Hiraoka, S., Saga, Y., Nagase, T., Ohara, O. and Koga, H.,  
TITLE Prediction of the Coding Sequences of Mouse Homologues of KIAA  
Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous  
cDNAs Identified by Screening of Terminal sequences of cDNA Clones  
Randomly Sampled from Size-Fractionated Libraries.  
JOURNAL Published Only in Database (2005)  
REFERENCE 2 (bases 1 to 2600)  
AUTHORS Okazaki, N., F.Kikuno, R., Nagase, T., Ohara, O. and Koga, H.

TITLE Direct Submission  
JOURNAL Submitted (22-FEB-2005) Hisashi Koga, Kazusa DNA Research  
Institute, Laboratory for Genome Informatics, 2-6-7  
Kazusa-Kametari, Kisarazu, Chiba, 292-0818, Japan  
(E-mail: mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)  
location/Qualifiers

FEATURES  
source  
1..2600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="mpj03374"  
/tissue type="embryonic tail"  
/note="vector:modified pbc sk+"  
<1..2600  
/gene="mKIAA4065"  
<1..2340  
/gene="mKIAA4065"  
/inference="non-experimental evidence, no additional  
details recorded"  
/note="CDS is predicted by in silico analysis. Start codon  
is not identified."  
/product="mKIAA4065 protein"  
/codon\_start=1  
/protein\_id="BAD90359.1"  
/db\_xref="gi:60360230"  
/translation="RRQMSPPGGGGGGGGGAGTRGAGRAGRRIIMGOTGKSKG  
/PCVRRKRSSEYRRLQKRFRADEVKTFNSRQKILERTTTLNDRKRIQEPV  
IMTSVSLRSTRCSVTSDDLDPFAQVPLKTLAAVAVSPIMVSPLOQNFVEDTV  
LHNIPYMGDEVLDODGTIELIKNTDQKXHGRCGFIINDEIFVEILVNAAGYND  
DDDDGDDPDERBKQDLBNRDKCTDPKRPADKIFEALISMPDKOTABELKX  
YKELTQGLPGLAPRECTPATIDGPNKRSVORBSLSPFTLFCRRGKQCFAPFPA  
TPNTYRKATETRLDNKPCGPOCYOGLBEARALTLARITTPKRPGRRGRLP  
NNSRSPTFTSVLESKDSDRAGFBRTGENDNDBEKKDPTSSSSANRRCQTP  
IKKPNLEPENWESGAESMFRVILGYDNPALILITKTCQVETFEVKES  
IIAPVTEVDPTPRKKRKRRLMAHCRKIQLKQGSNNHVNYPQCHPRQCDSS  
CPVIAQNFCEKFCOCSSSECONFPQRCYACQNTQCPCYLAVRECDPCLTCGA  
DHDSKRVSGKNGSIGIKHLLASPDVAGGIFLKDVGQNEFISEYCGEILSD  
EARRGVYDRTKNCSPLEFNINPVDATRKGNKIRPANSVNPNCAYKMMVNGDIR  
IGIFAKRAIQTSELFPDRTYSQADLKVIGIREMEIP"

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3756.00 Matches: 687  
Percent Similarity: 99.3% Conservative: 7  
Best Local Similarity: 98.3% Mismatches: 5  
Query Match: 98.6% Indels: 0  
DB: 6 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x AK220174 (1-2600)

Qy 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTyrPargLysArgValLys 20  
Db 100 ATGGGCCAGACTGGGAAGAAATCTGAGAAAGGACCGGTGTGGCGGAAGCGGTAAAA 159  
Qy 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
Db 160 TCAGAGTACATGACACGACGACGCTCAAGAGGTTCAAGAGAGCTGATGAAGTAAAGACT 219  
Qy 41 MetPheSerSerAsnArgGlnLysIleLeuGlnLysArgThrGluLysLeuAsnGlnGluTyr 60  
Db 220 ATGTTTAGTTCCATTCCTCGAAGAAATTTTGGAAAGAACTGAACCTTAAACCAAGATGG 279  
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyTyr 80  
Db 280 AAGCAGCGAGGATACAGCGCTGTGACATCATGACTTCTGTGAGCTCATTTGCCGGGACT 339  
Qy 81 ArgGluCysSerVal1ThrSerAspLeuAspPheProThrGlnVal1IleProLeuLysThr 100  
Db 340 AGGAGGTGTCACGACCAAGTGAAGTGAATTTTCCACACCAAGTCAATCCGTTAAAGACC 399  
Qy 101 LeuAsnAlaVal1AspValProIleMetTyrSerTyrSerProLeuGlnGlnAsnPhe 120

Db 400 CTGAATGACATCGCTCGGCTGCTATATGTAATCTTGGTCGCCCTTAACAGAAATTTT 459  
 Qy 121 MevValGluAspGluThrValLeuHISaenIleProTyrMetClyAspGluValLeuAsp 140  
 Db 460 ATGGTGAAGACAAACCTGTTTACATTAACATTCCTTAATATGGGGATGAAGTCTGGAGT 519  
 Qy 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
 Db 520 CAGATGACACTTCTTCTTAAGACACTAATTAATTAATGATGAGAAAGTGCATGGTGCAC 579  
 Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnIleLeuGlyGln 180  
 Db 580 AAGGAATGTGATTTATTAATGATGAATTTTGTGAGATGGTAAATGCTCTGGTCA 639  
 Qy 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluValGlyGlyGln 200  
 Db 640 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 699  
 Qy 201 LysAspLeuGluLysAspHisArgAspAspLysGluSerArgProAspGlyLysPheProSer 220  
 Db 700 AAGATCTAGAGATTAATGAGATGATTAAGAAATCTTGCCTCGAAATTTCTGCT 759  
 Qy 221 AspLysIlePheGluLysIleSerSerMetPheProAspLysGlyThrIleGluGluLeu 240  
 Db 760 GATTAATATTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGCAACCGAAGAACTG 819  
 Qy 241 LysGluLysTyrLysGluLeuThrGluGlnLeuProGlyValAlaLeuProProGluCys 260  
 Db 820 AAGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 879  
 Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
 Db 880 ACTCAAAATCATGATGACCAAAATGCCAAATCTGTTCAAGAGGAGAAAGCTTGCACTCA 939  
 Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisSerPheHis 300  
 Db 940 TTTCAATAGCTCTCTGTCGACGATGTTTAAATGATGATGATGATGATGATGATGATGAT 999  
 Qy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 Db 1000 GCAACACCCCAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1059  
 Qy 321 GlyProGlnCysTyrGlnHisIleLeuGluGlyValAlaLysGluPheAlaIleAlaLeuThrAla 340  
 Db 1060 GGAACACAGTGTATTCAGACATCTGAGAGAGCTTAAGAGTTCCTCTCTTATGCT 1119  
 Qy 341 GluArgIleLysThrProProLysArgProGlyLysArgArgArgLysArgLeuProAsn 360  
 Db 1120 GAGCGTATTAAGACACCACTTAACCCCAAGGGGGCCGAGAAAGAAAGACTTCCGAAAT 1179  
 Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 Db 1180 AACAGTACGACAGACCCGACACCCCAACATCAGTGTCTGATCAAAAGATACAGACACT 1239  
 Qy 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluGlyGlyGly 400  
 Db 1240 GACAG 1299  
 Qy 401 LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 Db 1300 AAAAGATGAGAGCTCCAGCTCTCTGAAGCAAAATTCCTGGTGTCAAAACCAATTAAGAG 1359  
 Qy 421 LysProAsnIleGluProProGluAsnValGluTyrSerGlyValAlaGluAlaSerMetPhe 440  
 Db 1360 AAGCCAAATATTAACCTCTGAAGATGTGAGATGAGTGTGTGAGAACCTCCATGTTT 1419  
 Qy 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1420 AAGATCTCTATTTGATCTTACTACATTAATCTTTTGTGCAATGTGCTTAATTTGGAC 1479  
 Qy 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 Db 1480 AAAACATGTAGACAGGTGTATGATTTAAGTCAAGAGATCCAGTATATCATACACTGTT 1539

Qy 481 ProIleGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
 Db 1540 CCACTGAGAGATGACACTCTCCAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1599  
 Qy 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 Db 1600 GCACTGTGAGAAATATCAATCTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659  
 Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 1660 CAACCTGTGACATTCACGAGAGCTTGTGTGACATTCCTGCTGTGTGTGTGTGTGTGT 1719  
 Qy 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 Db 1720 AATTTTGTGAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1779  
 Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys 580  
 Db 1780 CGGTGCAAGACCAATGCAACCAACAGTCCATGCTACTGCTGCTGCTGCTGCTGCTGCTGCT 1839  
 Qy 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisSerTyrAspSerLysAsnValSer 600  
 Db 1840 GACCTGACCTCTGTCTCACGCTGAGAGCTGCTGACATTTGGAGAGATTAATATGATCC 1899  
 Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysHisIleLeuLeuAlaProSerAsp 620  
 Db 1900 TGTAAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1959  
 Qy 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1960 GTGGAGGCTGGGAGATCTTATCAAGATCTGTACAGAAATTAATTAATTAATTAATTAAT 2019  
 Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 Db 2020 TACTGTGGGAGATTAATTTCTCAGATGACAGACAGAGAGAGAGAGAGAGAGAGAGAG 2079  
 Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
 Db 2080 TACATGTGAGCTTTCTGTCAACTTGAACATGATTTTGTGTGTGTGTGTGTGTGTGTGT 2139  
 Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 Db 2140 GGAACAAATTCGTTTGTCTATCATTCAGTAATCCAAATCTGATGCAAAAGTT 2196

RESULT 15  
 MMUS2951 2653 bp mRNA linear ROD 19-JUL-1996  
 LOCUS MMUS2951  
 DEFINITION Mus musculus putative transcriptional regulator mrx-1 mRNA,  
 complete cds.  
 ACCESSION U52951  
 VERSION U52951.1 GI:1279908  
 KEYWORDS vertebrate polycomb-group gene.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2653)  
 AUTHORS Robert O., Suess I., Closek T., Fuchs M. and Ullrich A.  
 TITLE Isolation and developmental expression analysis of Bmx-1, a novel  
 mouse Polycomb group gene  
 JOURNAL Mech. Dev. 55 (2), 171-184 (1996)  
 PUBMED 8861097  
 REFERENCE 2 (bases 1 to 2653)  
 AUTHORS Robert O.  
 TITLE Direct Submision  
 JOURNAL Submitted (28-MAR-1996) Department of Molecular Biology,  
 Massachusetts General Hospital, Wellman 8, Boston, MA 02114, USA  
 FEATURES  
 source  
 1..2653  
 /organism="Mus musculus"  
 /mol\_type="mRNA"





```
QY 541 AenpheCySGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGluCys 560
|||
Db 1755 AATTTTGGAAAAGTTTGTCAATGTAATTCAGAGTGTCAAAACCGCTTCTCGATGT 1814
|||
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580
|||
Db 1815 CGGTGCAAAAGCACAATGCAACACAAACAGTGTCCATGCTACCTGGCTGCCAGAGTGT 1874
|||
QY 581 AAPPProAspLeuCysLeuThrCysGluValAlaAspHisTTPAspSerLysAsnValSer 600
|||
Db 1875 GACCTGACCTGTGTCTCAGGTGAGCTGCTGACCATGGGACAGTAAATAATGTATCC 1934
|||
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysHisIleuLeuLeuAlaProSerAsp 620
|||
Db 1935 TGTAAAGAACTGTAGCATTTACGCGGGCTCTAATAAAGCACTTACTGCTGGCACCGTGTGAT 1994
|||
QY 621 ValAlaGlyTyrGlyTyrIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640
|||
Db 1995 GTGGCAGGCTGGGGGCACTTTTATCAAAAGATCCTGTACAGAAAATGAAATTCATCTCAGAA 2054
|||
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660
|||
Db 2055 TACTGTGGGAGATTATTTCTCAGAGTGAAGATGACAGAAAGGAAAGGTATGATATAA 2114
|||
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680
|||
Db 2115 TACATGTGCAGCTTGTGTCAACTTGAACAATGATTTGTGTGATGCCAACCGAANG 2174
|||
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699
|||
Db 2175 GGCACAAATAATCGTTTGTCTATCATTCAGTAAATCCAACTGCTATGCAAAAGTT 2231
|||
```

Search completed: August 13, 2006, 19:51:23  
Job time : 10526.6 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 13, 2006, 12:26:26 ; Search time 1085.99 Seconds  
(without alignments)  
6731.558 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699  
Perfect score: 3808  
Sequence: 1 MGQTKSEKBPVCMKRVK.....KGNKIRPANHVNPCYAKV 699

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues  
Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame.p2n.model -DEV=xlh  
-O=/abs/ABSSMB spool/US1077302/runat\_11082006\_140221\_16607/app\_query.fasta.1  
-DB=N Geneseg -OFT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0  
-UNIT=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=AB8804  
-USER=US1077302 @CGN 1.1 1518 @runat\_11082006\_140221\_16607 -NCPU=6 -ICPU=3  
-NO MAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THRBADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseg\_8:\*

1:	geneseqn1980s:*
2:	geneseqn1990s:*
3:	geneseqn2000s:*
4:	geneseqn2001s:*
5:	geneseqn2002s:*
6:	geneseqn2003s:*
7:	geneseqn2004s:*
8:	geneseqn2005s:*
9:	geneseqn2006s:*
10:	geneseqn2007s:*
11:	geneseqn2008s:*
12:	geneseqn2009s:*
13:	geneseqn2010s:*
14:	geneseqn2011s:*
15:	geneseqn2012s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB	ID
1	3808	100.0	2600 2 AAT43624
2	3808	100.0	2600 6 ABL65847
3	3808	100.0	2600 13 ADT90243

4	3808	100.0	2600 13 ADT89975	Adt89975 Human gen
5	3808	100.0	2600 13 ADT90094	Adt90094 Human gen
6	3808	100.0	2600 13 ADT90198	Adt90198 Human gen
7	3808	100.0	2711 12 ADQ22638	Adq22638 Human sof
8	3802	99.8	2558 10 ADG10641	Adg10641 Human STA
9	3802	99.8	2576 8 ABX76182	Abx76182 lung canc
10	3802	99.8	2576 11 ADK66988	Adk66988 Gene #78
11	3802	99.8	2576 11 ADN39391	Adn39391 Cancer/an
12	3802	99.8	2576 13 ADK25182	Adk25182 Breast ca
13	3802	99.8	2576 13 ADY61813	Ady61813 Human gen
14	3787.5	99.5	2253 12 ADP95922	Adp95922 Human E2H
15	3787.5	99.5	2695 10 ADP95922	Adp95922 Human E2H
16	3781	99.3	2512 10 ADD18523	Add18523 Human pro
17	3569.5	93.7	2476 10 ADB63038	Adb63038 Human STD
18	2863	75.2	1960 10 ADG10763	Adg10763 Human STA
19	2389.5	62.7	4185 14 ADZ61878	Adz61878 Murine Bz
20	2386.5	62.7	4606 14 AED17754	Aed17754 Fibrotic
21	2386.5	62.7	4640 14 ADX05847	Adx05847 Cyclin-de
22	2384.5	62.6	2594 10 ADK66987	Adk66987 Gene #77
23	2212.5	58.1	2522 8 ABX17535	Abx17535 Human CDN
24	1959.5	51.5	2448 4 AB111221	Ab111221 Drosophi1
25	1735.5	45.6	5289 4 AB111220	Ab111220 Drosophi1
26	1202	31.6	676 13 ADQ49105	Adq49105 Novel can
27	1138	29.9	714 6 ABL93013	Ab193013 Rat metas
28	1138	29.9	714 12 ADN07804	Adn07804 Human mam
29	875	23.0	22798 4 AAK89195	Aak89195 Human dig
30	875	23.0	22798 4 AAK81124	Aak81124 Human imm
31	776	20.4	485 9 ACH34755	Ach34755 Human end
32	723	19.0	3162 13 ADT16786	Adt16786 Plant CDN
33	719.5	18.9	2857 13 ADX30124	Adx30124 Plant ful
34	701.5	18.4	3030 6 AAK98824	Aak98824 CDNA of M
35	697.5	18.3	3061 12 ADM47813	Adm47813 Polynucle
36	696	18.3	5355 4 AAK89194	Aak89194 Human dig
37	693.5	18.2	1108 14 ADY65573	Ady65573 S. mansoni
38	679	17.8	3173 13 ADT16470	Adt16470 Plant CDN
39	667	17.5	3180 3 AAZ34943	Aaz34943 Corn poly
40	665	17.5	3148 6 AAK98823	Aak98823 CDNA of M
41	619	16.3	1948 13 ADX61599	Adx61599 Plant ful
42	617.5	16.2	2136 3 AAZ37895	Aaz37895 Fertilisa
43	617	16.2	2315 2 AAZ23136	Aaz23136 Arabidops
44	617	16.2	2315 4 AAF90504	Aaf90504 Arabidops
45	617	16.2	2799 8 ABZ68392	Abz68392 Nucleotid

## ALIGNMENTS

RESULT 1			
ID	AAT43624	standard; DNA; 2600 BP.	
XX	AAT43624;		
AC			
XX			
DT	25-MAR-2003 (revised)		
DT	05-MAY-1997 (first entry)		
XX			
DE	Chromatin regulator protein EZH2.		
XX			
KW	Chromatin; regulator; EZH1; EZH2; SUV39H; SET domain; gene therapy;		
KW	cancer; ss.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	90..2310	
FT		/*tag= a	
XX			
PN	W09635784-A2.		
XX			
PD	14-NOV-1996.		
XX			
PF	02-MAY-1996;	96WO-BP001818.	
XX			
PR	10-MAY-1995;	95DE-01016776.	

PA (BOEHR) BOEHRINGER INGELHEIM INT GMBH.  
XX  
XX Jenuwein T, Laible G;  
XX  
XX WPI; 1996-518672/51.  
DR P-PSDB; AAM05260.  
XX  
XX New DNA encoding chromatin regulator protein with SET domain - and  
PT related vectors, transformed cells, proteins and antibodies, for  
PT diagnosis and treatment of cancer.  
PT  
XX  
XX Claim 1; Fig 6; 38pp; German.  
PS  
XX  
XX The DNA was isolated by screening a human B cell cDNA library with mixed  
CC  
CC Drosophila DNA probes based on the conserved SET domains in R(z) and  
CC Su(Var)3-9. The DNA, and its products, are useful in therapy (esp. gene  
CC therapy) and diagnosis of human diseases that involve deregulated  
CC chromatin-regulator genes having a SET domain, esp. cancer. (Updated on  
CC 25-MAR-2003 to correct PR field.)

**SQ** Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	2,38e-282
Score:	3608.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	2
	Gaps: 0
	Mismatches: 0
	Conservative: 0
	Matches: 699
	Length: 2600

US-10-773-302-2 COPY 1 699 (1-699) X AAT43624 (1-2600)

OY	1	MetGlyGlnThrGlyLysLeuYysSerGluYysGlyProValCysTrpArgLysArgValLys	20
Db	90	ATGGGCGCAGACTGGGAGAAATCTGAGAAAGGACCGATTGTTGGGGAAGCGGTAA	149
OY	21	SerGluYrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer	40
Db	150	TCAGGTACATCCGACTGACGACGCTCAAGAGTTCAACGAGCTGATGAATGAAGAAT	209
OY	41	MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGluIleLeuAsnGlnGluTrp	60
Db	210	ATGTTTAGTTCCAATCGTCAGAAAATTTTGGAAAGAACCGAAATCTTAAACCAAGATGG	269
OY	61	LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr	80
Db	270	AAACGCGGAAGATACGCTTGCCATCTGATCTGTGAGCTCATTTGGCGGGAGCT	329
OY	81	ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysTrp	100
Db	330	AGGGAGTGTCCGTCGACCACTGACTTGGATTTTCCAACCAAGTATCCATTAAAGACT	389
OY	101	IleuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnIleAsnPhe	120
Db	390	CTGAATGACGTTGCTTCAGTACCACCAATATGATTTCTTGCTCCCTACAGCAGATTTT	449
OY	121	MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAsp	140
Db	450	ATGGTGGAAGAGAACTGTTTAACTPAACATCTCTTAATATGGGAATGAAGTTTAAAGT	509
OY	141	GlnAspGlyThrPheIleGlnIleuLysIleLysAsnTyrAspGlyLysValHisGlyAsp	160
Db	510	CAGGATGGTACTTCTCTTAAGAACTATATAAAAATTAATGATGGGAAAGTACCGGGAT	569
OY	161	ArgGluCysGlyPheIleAsnAspGluIlePheValGlnLeuValAsnAlaLeuGlyGln	180
Db	570	AAGAGATGGGGTTTAAATGATGAATTTTGTGGAAGTGGTGAATGCTTGGTCA	629
OY	181	TyrAsnAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnGlyLysGln	200
Db	630	TATTAATGATGATGACATATGATGATGATGAGCGAGTCTGTGAAGAAAGAAAGAACG	669

QY	201	LYSAERLEUGLIASHPHISATGAAPARLYSGIUSEZARXPROBROKTRGYSPHEPROSER	220
DB	690	AAACATCTCGAGGATCCGAGATGATAAAGAAAGCCGACCTCGGAAATTTCTCTT	749
QY	221	ASPRYLITLPEHGLIUALIIESERSEMEXPHROAPRYSGIYLTHRALAGIUGIULEU	240
DB	750	GATTAATTTTTTGAAAGCAATTTCTCAATCTTTCCAGATPAAGGGCACAGCAAGAACTA	809
QY	241	LYSGIULYSTRYIUGIULEUTHRGUGINGIULEUPROGIYALALEUPROPROGIUCYB	260
DB	810	AAGGAAAAATTAAGAACTCACCGAAACACACTCCACGCGCACTTCTCTGTGAATGT	869
QY	261	THYPROASNTLEAPGIYPROASNTALYSESVALGIRARGIUGINSEITLSEHSER	280
DB	870	ACCCCAACATAGATGACCAAAAGCTTAAATCTGTTCAGAGAGAGCAACCTTACACTCC	929
QY	281	PHENITHIRLEUPHECYBARGARCYBPHELYSTRYBZARCYBPHLEUHIAPROPHENIS	300
DB	930	TTTCATACGCTTTCTGTAGGCGATGTTTTAAATATACCTGCTTCTCAACTCTTTTCAT	989
QY	301	ALATHYPROASNTHTYTRYLSARGLYSASNTHTGIUTHRALALEUPRANLYSPROCYB	320
DB	990	GCAACACCCCAACACTTATTAAGCGGAGAAACACAGAAACACTCTAGCAACAAACCTTGT	1049
QY	321	GLYPROGICNYSTRYGINHISLEUGIUGIYALALEGIULPHELAIALALEUTHRALA	340
DB	1050	GGACCAACGTGTATCCAGCATTTGGAGGGAGCAAGAGTTTCTGTCTCACCGCT	1109
QY	341	GIULYRGITLHYSTRHTPROPOLYBAXYPROGIYGIYRGATRGIGIYATGLEUPROASN	360
DB	1110	GAGCGATATAGACCCCAACCAACCTCCAGAGCGCGCAAGAGAGACGGCTTCCCAAT	1169
QY	361	ASNSERSEARXPROSETRHTPROTHIRLEASNVALLEUGISERYBAPHTHARPSER	380
DB	1170	AAACGTAGCAAGCCACAGACCCCAACCATTAATGTCTGGAATCAAGATATACAGACAGT	1229
QY	381	ASPARGIUAIAGIYTHRGUITHGLYGLYGLIULASNSAPRYSGIUGIUGIULYB	400
DB	1230	GATGGAGAGACGAGGACTGAACCGGGGAGAGCAATATTAAGAAAGAAAGAGAG	1289
QY	401	LYSAPRGUITHRSESERSESERSESGIUALASNSERXRGYSGINTHTPROILEYSEMET	420
DB	1290	AAAGATGAATCTTGAGTCTCTCTGAAGCAATTTCTGGTGTCAACACCAATPAAGATG	1349
QY	421	LYSPROASNTLEGIUPROPROGIULASNTVALGIUTPSEGIYALAGIUALASERMEPHE	440
DB	1350	AAAGCAATATTTGAACCTCCGAGAAATGTGAGTGGTGTGTAAGCTCAATGTTT	1409
QY	441	ATGVALLEUITLEGIYTHTRYTRYTASPSAPNHEYSALIALEIAAAGLEUITLEGIYTHR	460
DB	1410	AGAGTCCATTTGGGCACTTACTATGACAAATTTCTGTCCCATTTGTAGATTAAATGGGACC	1469
QY	461	LYSTHTCYBARGGINVALTYRGUINPHARYVALLYSGIUSERSEITLIEALAPROLA	480
DB	1470	AAAAATGTAAACAGGTGTATGAGTTTAAAGATCAAAAGATTTAGCATCAATGCTCCAGCT	1529
QY	481	PROALAGIUAERVALASPTHTPROBROKTRGYLYBYLYBARGLYBHIARGLEUTPDALA	500
DB	1530	CCCGGTGAGATGTGGATACCTCTCTCCAAAGAAAAAGAAAGAAACACCGGTTGTGGGCT	1589
QY	501	ALAHISCYBARGLYBIEGINLEUULYLYBAPGLYSESERASNTHEVALTYRASNTRYR	520
DB	1590	GCACACTCGAAGAAAGATCACACTGAAGAAAGACGGCTCTCTTAACATGTTTACACTAT	1649
QY	521	GINPROCYBAPRHIAPROARGGILPROCYBAPSESESECYBPROCYBVALIIEALGIN	540
DB	1650	CAACCTTGATCATCCACGGCAGCTTGTATACAGTTGTGCCCCCTTGTGTATGCAAMA	1709
QY	541	ASNPHECYSGIULYBPHCYBGLINCYSESESESGIUCYBGLINASATRPHETROGIYCS	560
DB	1710	AAATTTTGTGAAAAAGTTTTTGTCAATGTATGTTCAAGATGTCAAAACCGCTTTCCGGAGATGC	1769
QY	561	ATGCGYLYBVALAGINCYBASNTHRYBGINCYBAPROCYSTRILEUALALEVALARGIUCYB	580

```
DB 1770 CGCTGCAAGACAGTGCACACCAAGTCCGCTGCTACTGCTCCGAGAGTCT 1829
QY 581 AAPPProAspLeuCysLeuThrcysGlyValaalaasphsITPAspSerLyAsnValSer 600
DB 1830 GACCTGACCTCTGCTTACTGTGTGAGCCGCTGACCATTTGGAGACAGTAAATGTGTCC 1889
QY 601 CysLyAsnCysSerIleGlnArgGlySerLyAsnIleLeuLeuLapProSerAsp 620
DB 1890 TCGAAGAACTGCAAGTATTCAGCGGGGCTCAAAAAGCATCTATGTGCGACCATCTGAC 1949
QY 621 ValAlaGlyTrpGlyIlePheIleLyAsnProValGlnLyAsnGlnPheIleSerGlu 640
DB 1950 GTGGCAGGCTGGGGGATTTTATCAAGATCCTGTGCGAAGAAAATGATTCATCTCAGA 2009
QY 641 TTYCysGlyGlnIleIleSerGlnAspGluAlaAspArgArgGlyLyValIlyAspLy 660
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGGGAAAGTATGATTA 2069
QY 661 TTYMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLy 680
DB 2070 TACATGTGAGCTTTCTGTCTCACTTGAAACAATGATTTGTGTGATGCAACCGCAG 2129
QY 681 GLyAsnLyIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLyVal 699
DB 2130 GGTAAACAAATTCGTTTCCAAATCATTCGTAATCCAAACTGCTATGCAAAAGTT 2186

RESULT 2
ABL65847
ID ABL65847 standard; DNA; 2600 BP.
XX
AC ABL65847;
XX
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4184.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
```

```
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 03-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 4184; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'
CC tumour
XX
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;
XX
XX

Alignment Scores:
Pred. No.: 2,38e-282 Length: 2600
Score: 3808.00 Matches: 699
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-10-773-302-2_COPY_1_699 (1-699) x ABL65847 (1-2600)
QY 1 MetGlyInThrGlyLyLyAsnSerGlnLyAspProValCysTrpArgLyAsnValLy 20
DB 90 ATGGGCCAGACTGGGAAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAGGTATAAA 149
QY 21 SerGlyTyrMetArgLeuArgGlnLeuLyAsnArgPheArgAlaAspGlnValLySer 40
DB 150 TCAGAGTACATGACGACTGACAGCAGCTCAAGAGGTTCAACACGAGCTGATGAAGTAAAGAGT 209
QY 41 MetPheSerAsnArgGlnLyIleLeuGlnArgThrGlnIleLeuAsnGlnIleTrp 60
```

D	210	ATGTTAGTTCGAATCGTCAGAAAATTTGGAAAAGACGGAATCTTAACCAAGATCG	265
Q	61	LygsluArgArgllleglnProValHlsileuthSerValSerSerleuArglyThr	80
D	270	AAACGAGGAAAGATACAGCTGTGCATCTGACTTGTGAGGTCACTTCGGGGGAGCT	329
Q	81	ArgGluCySerSerValThrSerAspLeuAspPheProThrGlnValIleProleuLyThr	100
D	330	AGGAGGTGTTCGGGACCAAGTGAACCTTATTTCCACACAAAGTCATCCCATTAAGACT	389
Q	101	LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProleuGlnAlaAspHe	120
D	390	CTGAATGCAAGTTGCTCAGTACCCATATGATTTCTTGCTCCCTACAGAGAAATTTT	449
Q	121	MetValGluAspGluThrValIleuHlsAsnIleProTyrMetGlyAspGluValLeuAsp	140
D	450	ATGCGTGAAGATGAAACCTGTTTACATACATCTCTTATATGGAGATGAAGTTTATGAT	509
Q	141	GlnAspGlyThrPheIleGluGluLeuIleLyAsnTyrAspGlyLyValHlsGlyAsp	160
D	510	CAGATGGTACTTTCATTTGAAGAACTAATAAAAAATTTGATGGGAAAGTACACGGGAT	569
Q	161	ArgGluCySerGlyPheIleAsnAspGluIlePheValGluLeuValAlaAsnAlaGln	180
D	570	AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGTGAATGCCCTTGTGTCA	629
Q	181	TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluGluGluGln	200
D	630	TATATGATGATGACGATGATGATGATGACGATCTCGAAGAAAGAAAGAAAGAGAG	689
Q	201	LyAspLeuGluAspPheIleArgAspAspGlySerAlaGlyProProArgLyPheProSer	220
D	690	AAAGATCTGGAGGATCACCGAGATGATTAAGAAACCCGCCACCTCGGAAATTTCTTCT	749
Q	221	AspLySIIlePheGluAlaIleSerSerMetPheProAspLySIIleThrAlGluGluLeu	240
D	750	GATTAATTTTGAAGCCATTTCTCATATGTTTCCAGTAAAGGACACGCGAAGAACTA	809
Q	241	LySIIleGlySIIleGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGluCys	260
D	810	AAAGAAAATATAAGAACTACACGACAGACGCTCCAGGGGACCTTCCTCGAATGT	869
Q	261	ThrProAsnIleAspGlyProAsnAlaLySerValGlnAspGluGlnSerLeuHlsSer	280
D	870	ACCCCAACATAGATGACCAAAATGCTTAATCTGTTACAGAGAGCAAAAGCTTACACTCC	929
Q	281	PheIleThrLeuPheCysArgArgCysPheLySIIleThrAspCysPheLeuHlsProPheHls	300
D	930	TTTCATAGCTTTTCTGTAGGGAGATGTTTAAATATGACTGCTTCTTCATCTTTTCAT	989
Q	301	AlaThrProAsnThrTyrLyAspGlyAspAsnThrGluThrAlaLeuAspAsnLyAspCys	320
D	990	GCAACACCCCAACACTTATTAAGGGAGAAACACAGAAACAGCTCTAGACAAACCTTGT	1049
Q	321	GlyProGlnGlyTyrGlnHlsleuGluGlyAlaLySIIlePheAlaAlaAlaLeuThrAla	340
D	1050	GGACCAACAGTGTTCACACATTTTGAAGGAGCAAAAGAGTTTGCCTGCTCAACCCCT	1109
Q	341	GluArgIleLyThrProProLyAspProGlyGlyAspAspArgGlyAlaArgLeuProAsn	360
D	1110	GAAGCGATTAAGACCCCAACCAACGTCACAGAGCCCGCAAGAGAGCGGCTTCCCAT	1169
Q	361	AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLyAspThrAspSer	380
D	1170	AAACGTACAGGCCCAACAGACCCCAACATTAATGCTGATCAAGATACAGACAGT	1229
Q	381	AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLySIIleGluGluGln	400
D	1230	GATGGGAGAGCAAGGATCGAAACGGGGGAGAGAACTATGATTAAGAAAGAAAGAGAG	1289
Q	401	LyAspAspGluThrSerSerSerSerGluAlaAsnSerArgGlnThrProIleLyMet	420
D	1290	AAAGATGAACCTTCAGACTCTCTGAAGCAAAATCTCGGTGTCAACACCAATTAAGATG	1349

---

Q	421	LyProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe	440
D	1350	AAAGCAATATATGAACCTCTGAGAAATGTGAGAGGTGTGTGTAAGCTTCATATGTTT	1409
Q	441	ArgValLeuIleGlyThrTyrTyrAspAspPheCysAlaIleAlaArgLeuIleGlyThr	460
D	1410	AGAGTCTCATGGCACTTACTATGACAAATTTCTGTCCATTTGCTAGATTAAATGGAGCC	1469
Q	461	LyThrCysArgGlnValTyrGluPheArgValIleGluSerSerIleIleAlaProAla	480
D	1470	AAACATGTACAGAGGTATGATGAGTTAAGTCAAAAGATCAATAGCTCCAGCT	1529
Q	481	ProAlaGluAspValAspThrProProArgLyLySIIleLyAspGlyHlsAlaGluThrAla	500
D	1530	CCCCCTGAGATGTGGATACCTCCCAAGGAAAGAAAGAGAAACACCGGTTGTGGCT	1589
Q	501	AlaHlsCysArgLySIIleGlnLeuLyAspGlySerSerAsnHlsValTyrAsnTyr	520
D	1590	GCACATGCAAGAAAGATACAGCTGAAGAAAGAGCGCTCTCTAACATGTTTACACTAT	1649
Q	521	GlnProCysAspHlsProArgGlnProCysAspSerSerCysProCysValIleAlaGln	540
D	1650	CAACCTGTATCATCCACGGCAGCTTGTGACAGTTCTGTGCCCTTGTGTATGACAA	1709
Q	541	AsnPheCysGluLySIIlePheCysArgGlnCysSerSerGlyCysGlnAsnAspPheProGlyCys	560
D	1710	AATTTTGTGTAAGAGTTTGTGCATGTGATTCAGAGTGCATAAACCCGCTTCCGGAGTC	1769
Q	561	ArgCysLySIIleGlnCysAsnThrLySIIleCysProCysTyrLeuAlaValArgGluCys	580
D	1770	CGCTGAAAGACAGTTCACACCAACAGTCCCGCTGCTACCTGCTGCCAGATGT	1829
Q	581	AspProAspLeuCysLeuThrCysGlyAlaAlaAspHlsThrAspSerLyAsnValSer	600
D	1830	GAACCTGACCTGTCTTACTTGTGAGCCGCTGACCATTTGGACATTAATAATGTGTCC	1889
Q	601	CysLyAsnCysSerIleGlnArgGlySerLySIIleSleuLeuLeuAlaProSerAsp	620
D	1890	TGCAGAACTGCAGTATTCAGCGGGGCTCCAAAAGATCTATTTGCTGGCACCACTTCAC	1949
Q	621	ValAlaGlyTyrGlyIlePheIleLyAspProValGlnLyAsnGluPheIleSerGlu	640
D	1950	GTGGCAGCGGGGATTTTATCAAAAGATCTGTGGCAAGAAATGAATTCATCTCAGAA	2009
Q	641	TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLySIIleTyrAspLyS	660
D	2010	TACTGTGAGAGATTTATTTCTCAAGATGAAGCTGACAGAGAGGAAAGTATGATAAA	2069
Q	661	TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValaAspAlaThrArgLyS	680
D	2070	TACATGTGACGCTTCTGTTCACCTTGAAACATATTTTGTGGATGTGCAACCGGAGG	2129
Q	681	GlyAsnLySIIleArgPheAlaAsnHlsSerValAsnProAsnCysTyrAlaLyVal	699
D	2130	GGTAACAAATCTGTTTGCATATCATTTCCGTAATCCAAACTGCTATGCAAAAGTT	2186

RESULT 3  
ADT90243  
ID ADT90243 standard; DNA; 2600 BP.  
AC ADT90243;  
NC  
DT 13-JAN-2005 (first entry)  
DE Human gene over-expressed in a cancerous cell, SEQ ID 356.  
XX cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
XX cytotoxic; gene therapy; gene; ds.  
XX Homo sapiens.  
OS  
XX  
FN WO2004091548-A2.





QY 481 ProIaGluApValAspThrProProArgLysLysArgLysHisArgLeuTyrAla 500  
DB 1530 CCGCTGAGATGATGATTAAGTCTCTCAAGAGAAAAGAGAGAACACCGGTTGTGGCT 1589  
QY 501 AlAhicGySaAglYsIleGlnLeuLysLysAspLysSerSerAsnHisValTyrAsnTyr 520  
DB 1590 GCACACTGCAAGAAAGATACAGCTGAAAAGACGGCTCTCTTAACATGTTTACACATAT 1649  
QY 521 GlnProCyAspHisProArgGlnProCyAspSerSerCyProCyValIleAlaGln 540  
DB 1650 CACCTGTATCATCTCCAGCGACCTTGTACAGTCTGTCCCTTGTGTATGACAA 1709  
QY 541 AsnProCyGluLysLysLysGlnCySerSerSerGlnCyGlnAsnArgPheProGlyCy 560  
DB 1710 AATTTTGTGAAAAGTTTGTCAATGTATGTTCAAGTGTCAAAACCGCTTCCGGGATGC 1769  
QY 561 ArgCyLysValAlaGlnCyAsnThrLysGlnCyProCyGlyTyrLeuAlaValArgGlnCy 580  
DB 1770 CGCTGCAGAAAGCAAGTGCACACCAAGCAAGTGCCTGTACTGTGCTGCAGAGATGT 1829  
QY 581 AspProAspLeuCySerLeuThrCyGlyValAlaAspHisTyrPAspSerLysAsnValSer 600  
DB 1830 GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCATTTGGACAGTAAATGTGTCC 1889  
QY 601 CyLysAsnCySerIleGlnArgGlySerLysLysLysLeuLeuAlaProSerAsp 620  
DB 1890 TGCAAGAACTGACAGTATTCAGCGGGCTCCAAAAGCATCTATTGCTGCACATCTGAC 1949  
QY 621 ValAlaGlyTyrGlyLysPheLysAspProValGlnLysAsnGlnPheIleSerGln 640  
DB 1950 GTGCAGAGCTGGGGATTTTATCAAAAGATCTGTGCAGAAAATGAATTCATCTCAGAA 2009  
QY 641 TyrCyGlyGlnLysLysSerGlnAspGlnAlaAspArgGlyLysValTyrAspLys 660  
DB 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAAAGGAAAGTGTATGATTA 2069  
QY 661 TyrMetCySerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
DB 2070 TACATGTGACACTTCTGTCTCAACTGAACATGATTTTGTGTGATGCAACCGCGCAG 2129  
QY 681 GlnLysLysLysLysPheAlaAsnHisSerValAsnProAsnCyTyrAlaLysVal 699  
DB 2130 GGTAACAAATTCGTTTGGCAATCATTCGGTAATCCAAACCTGATGCAAAAGTT 2186

RESULT 4  
ADT89975 standard; DNA; 2600 BP.  
ID ADT89975;  
AC ADT89975;  
DT 13-JAN-2005 (first entry)  
DB Human gene over-expressed in a cancerous cell, SEQ ID 88.  
XX cancer; detection; pre-cancerous; antineoplastic; cancer therapy;  
XX cytosolic; gene therapy; gene; de.  
XX Homo sapiens.  
XX WO2004091548-A2.  
XX 28-OCT-2004.  
XX 15-APR-2004; 2004WO-US009289.  
XX 15-APR-2003; 2003US-0462895P.  
XX (AVAL- ) AVALON PHARM INC.  
XX Strovrel JW, Cain CB, Horrigan SK, Augustus M,  
XX WPI; 2004-784489/77.

PT Diagnosing cancer in a mammal comprises determining amplification of  
PT specific genes, e.g., serine/threonine kinase 6, in the genome of a  
PT mammal.  
XX  
PS Claim 3; SEQ ID NO 88; 92pp; English.  
XX  
CC The invention relates to a novel method for diagnosing cancer in a  
CC mammal. The method comprises determining amplification of specific genes,  
CC given in the specification, in the genome of a mammal. The invention  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
CC human patient; identifying an anti-neoplastic agent; determining the  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytotoxic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 2,38e-282 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 13 Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x ADT89975 (1-2600)  
QY 1 MetGlyGlnThrGlyLysLysSerGlnLysGlyProValCyStrPAspGlyAspValLys 20  
DB 90 ATGGGCCAGACTGGGAAGAAATCTGAGAGAGGACCAAGTTGTGGCGGAAGCGGTAA 149  
QY 21 SerGlnLysPheLysArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
DB 150 TCAGAGTACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 209  
QY 41 MetPheSerSerAsnArgGlnLysLysLysLysLysLysLysLysLysLysLysLys 60  
DB 210 ATGTTTACTTCATCTCTCAGAAATTTTGGAAAGAACGGAATCTTAACCAAGATGG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgLysThr 80  
DB 270 AAACAGCGAAGAAATACAGCTGTGCACATCTCTGAGCTCATTTGCGGGACT 329  
QY 81 ArgGlnCySerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 330 AGGAGGTTCGAGCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTyrSerProLeuGlnAsnPhe 120  
DB 390 CTGAATCAGTGTCTTCAGTACCAATATGATTTCTGTCTCCCTACAGCAGAAATTT 449  
QY 121 MetValGlnAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAsp 140  
DB 450 ATGCTGAGAGATGAACGTTTTCATTAACATCTCTTAATGAGAGATGAAGTTTATGAT 509  
QY 141 GlnAspGlyThrPheLysLysLysLysLysLysLysLysLysLysLysLysLysLys 160  
DB 510 CAGGATGATCTTTCATTTGAAGAACTAATTAATAATTAATTAATTAATTAATTAATTA 569  
QY 161 ArgGlnCyGlyPheLysAsnAspGlnLysPheValGlnLeuValAlaAsnAlaLeuGln 180  
DB 570 AGAAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGAGTGTGAGTGTGAGT 629

[illegible]

Db	1710	AAATTTTGTGAAAAGTTTGTGCATGTATGTTACAGAGTGTCAAAACCGCTTCCGGATGC	1769		
Qy	561	ArgCysIysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys	580		
Db	1770	CGTGCAAAAGCAGGTGCAACACCAAGCAGTGCCTGGTCACTGGCTGTCCGAGAGTGT	1829		
Qy	561	AspProAspLeuCysLeuThrCysGlyValAlaAspHisThrAspSerLysAsnValSer	600		
Db	1830	GACCTTGACCTCTGTCTTACTGTGTGAGCCGCTGACCATTTGGAGCAGTAAAAATGTCTCC	1889		
Qy	601	CysIysAsnCysSerIleGlnArgGlySerLysHisLeuLeuAlaProSerAsp	620		
Db	1890	TGCAGAGACTGCATATTCACGGGGGCTCCAAAAGCATTTATTTGCTGGACATCTGCAC	1949		
Qy	621	ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu	640		
Db	1950	GTGGCAGGCTGGGGGATTTTATTCAAAGATCCTGTGCAGAAAATGAAATTCATCTCAGAA	2009		
Qy	641	TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys	660		
Db	2010	TACTGTGAGAGATTATTTCTCAGATGAAGCTGCACGAGAGGGGAAAGTGTATGATMAA	2069		
Qy	661	TyrlMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys	680		
Db	2070	TACATGTGCAGCTTTCGTGTCACTTAAACAATGATTTTGTGTGATGCAACCCGCAAG	2129		
Qy	681	GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal	699		
Db	2130	GGTAACAAAAATGCTTTTGGCAAAATCATTCGGTAAATCAAAAGCTATGCAAAAGTT	2186		
RESULT 5					
ID	ADT90094	ADT90094 standard; DNA; 2600 BP.			
XX	AC	ADT90094;			
XX	DT	13-JAN-2005 (first entry)			
XX	DE	Human gene over-expressed in a cancerous cell. SEQ ID 207.			
XX	KW	cancer; detection; pre-cancerous; antineoplastic; cancer therapy;			
XX	KX	cyrostatic; gene therapy; gene; ds.			
XX	OS	Homo sapiens.			
XX	PN	MO2004091548-A2.			
XX	PD	28-OCT-2004.			
XX	PF	15-APR-2004; 2004WO-US009289.			
XX	PR	15-APR-2003; 2003US-0462895P.			
XX	PA	(AVAIL-) AVALON PHARM INC.			
XX	PI	Strovel JW, Cain CB, Horriigan SK, Augustus M;			
XX	DR	WPI; 2004-784489/77.			
XX	PT	Diagnosing cancer in a mammal comprises determining amplification of			
XX	PT	specific genes, e.g., serine/threonine kinase 6, in the genome of a			
XX	PS	mammal.			
XX	CC	Claim 3; SEQ ID NO 207; 92pp; English.			
CC	The invention relates to a novel method for diagnosing cancer in a				
CC	mammal. The method comprises determining amplification of specific genes,				
CC	given in the specification, in the genome of a mammal. The invention				
CC	further comprises methods for: diagnosing cancer or a pre-cancerous				
CC	condition in a mammal; inhibiting cancer or a pre-cancerous condition in				
CC	a mammalian cell; identifying an agent having therapeutic activity in a				
CC	human patient; identifying an anti-neoplastic agent; determining the				

CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
CC patient; determining the likelihood of success of cancer therapy in a  
CC patient; producing test data with respect to the anti-neoplastic activity  
CC of a compound; and determining the progress of a treatment for cancer in  
CC a patient following commencement of a cancer treatment on the patient.  
CC The methods and compositions of the invention have cytostatic activity  
CC and may be used in gene therapy. The method is useful in diagnosing or  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
CC diagnosis method of the invention.

XX Sequence 2600 BP, 818 A, 525 C, 630 G, 627 T, 0 U, 0 Other;

## Alignment Scores:

Pred. No.:	2,386-282	Length:	2600
Score:	3808.00	Matches:	699
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	13	Gaps:	0

US-10-773-302-2\_COPY\_1\_699 (1-699) x ADP90094 (1-2600)

QY 1 MetGlyInThrGlyLeuLeuSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
DB 90 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGAGCCAGTTTGTCGGGAGACGCTGTAATAA 149  
QY 21 SerGlyTrpMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGlnValLysSer 40  
DB 150 TCAAGATCATGCGACGTAGACAGCTCAAGAGTTCAACAGACGATGAAAGTAAAGAGT 209  
QY 41 MetPheSerSerAspArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnGluTrp 60  
DB 210 ATGTTTAGTTCATTCGTCAAGAAATTTTGGAAAGAACGAAATTTTAAACAGAAAGCG 269  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 270 AAACGCGAAGATGACAGCTGCTGCAATCCGACTTCGTGAGCTCATTCGCGGAGCT 329  
QY 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 330 AGGAGTGTTCGGTGACCAAGTCACTTGATTTTCCAAACAGATCATCCATTAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnAsnPhe 120  
DB 390 CTGAATGCACTTCTCACTACCAATTAATGTTTCTGTCCTCCCTACAGAGAAATTTT 449  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 450 ATGGGGAAGATGAATCTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTAAAT 509  
QY 141 GlnAspGlyThrPheIleGlnLysLeuLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 510 CAGGATGTACTTCAATTAAGAACTAATTAATAAATTAATGATGGGAAAGTACACGGGAT 569  
QY 161 ArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlyGln 180  
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGCTGAATGCCCTTGCTCA 629  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGlnArgGlnLysGln 200  
DB 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689  
QY 201 LysAspArgLeuGlnAspHisArgAspAspLysGlnAspArgProGlnArgLysPheProSer 220  
DB 690 AAAAGTCTGGAGAGATCACCGAGATGATTAAGAAACCGCCCACTCGGAAATTTCTCTCT 749  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLysLeu 240  
DB 750 GATTAATTTTGGAGCCATTTCTCATGTTTCCAGATTAAGGACACAGCAAGAAACATA 809  
QY 241 LysGlnLysTyrLysGlnLeuThrGlnGlnLeuProGlyAlaLeuProProGlnLys 260

DB 810 AAGGAAAAATTAAGAACTACCGAAGACAGCTCCAGGCGACTTCTCTGAATGT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnIleSerLeuHisSer 280  
DB 870 ACCCCCAACATATAGTGAACCAATGCTTAATCTGTTCCAGAGAGCAAACTTACATCC 929  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 930 TTTCATACGCTTTTCTGTAGGCGATGTTTAATATGATGATGATGATGATGATGATGAT 989  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
DB 990 GCACACCCCAACATTAATTAAGCGGAGAACACAGAAACAGCTTCAACACAACTTGT 1049  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGlnGlyValLysGlnPheAlaAlaIleuThrAla 340  
DB 1050 GACCAACAGTGTTCACAGCATTTTGAGAGGACAAAGAGATTTGCTGCTTCACCGCT 1109  
QY 341 GlnArgIleLysThrProProLysArgProGlyValArgArgArgGlyArgLeuProAsn 360  
DB 1110 GAGCGATTAAGAACCCCAACAAACGTCCAGAGGCGCGAGAGAGAGCGCTTCCCAAT 1169  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1170 AACATGACAGGCGCCAGACACCCCACTTAATGTGCTGAATCAAGATACAGACAT 1229  
QY 381 AspArgGlnAlaGlyThrGlnThrGlyGlnLysAsnAspLysGlnGlnLys 400  
DB 1230 GATAGGAGAGGAGGACTGAAGACGGGGGAGAGAACATGATTAAGAAAGAAAGAGAG 1289  
QY 401 LysAspGluThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1290 AAAGATGAATCTTGACCTCCTCGAAGCAAAATCTCGGTCTCAACACCAATTAAGATG 1349  
QY 421 LysProAsnIleGlnProProGlnAsnValGluThrSerGlyAlaGlnAlaSerMetPhe 440  
DB 1350 AAGCCAAATATTGAATCTCTGAGAAATGTGAGAGGTGTGTGAAGCTTCATATTT 1409  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1410 AGAGTCTCATGTGCACTTACTACTAGACATTTCTGTGCTCATTGCTAGTTAATGGAGCC 1469  
QY 461 LysThrCysArgGlnValTyrGlnPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB 1470 AAAACATGATGACAGGTGATGAGTTTGAAGTCAAAACATGACATCATGCTCCAGCT 1529  
QY 481 ProIleGlnAspValAspThrProProArgLysLysAspArgLysHisArgLeuThrAla 500  
DB 1530 CCGCTGAGAGTGTGATACTCTCCAAAGGAAAGAGAGAAACACCGGTGTGGGCT 1589  
QY 501 AlaHisCysArgLysIleGlnLeuLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1590 GCACATGCGAAGAAAGATACAGCTGAAGAAAGAGCGCTCTCTTAACATGTTTACAAT 1649  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1650 CAACCTGTGATCTTCAACCGGACGCTTGAGACGTTGCTGCTTGTGTATGACAA 1709  
QY 541 AsnPheCysGlyLysPheCysGlnCysSerSerSerGlnCysGlnAsnAspPheProGlyCys 560  
DB 1710 AATTTTGTGAAAGTTTGTCAATGTATGATCAAGTGTCAAAACCGCTTCCGGGATGC 1769  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGlyCys 580  
DB 1770 CGCTGCAAGACAGTGCAGACACCAAGAGTGGCGCGTGTACCTGGCTGTCCAGAGTGT 1829  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1830 GACCTGACCTCTGTCTTACTTGTGAGCGGCTGACCATTTGGGACAGTAAATAATGTGTCC 1889  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAsp 620

DB 1890 TCAGAAAGCTGCATATTCAGCGGCGCTCCAAAAGCATCTATTCGTGGCACCACATCTGAC 1949  
QY 621 VAlAlAGlYTrGClYlIePheIlleYsAPProValGlnLYsAnglYlIeSerglY 640  
DB 1950 GTGGCAGGCTGGGGATTTTATCAAAAGATCCTGTCGCAAAAATGAAATTCATCTCAGAA 2009  
QY 641 TTrCYsGlyGlnlIeIleSerglYlIeSerglYlIeSerglYlIeSerglYlIeSerglY 660  
DB 2010 TACTGTGAGAGATTTTCTCTCAAGTGAAGCTGACAGAGAGGAAAGTGTATGATATA 2069  
QY 661 TTrMetCYsSerPheLeuPheLeuPheLeuPheValValIAspAlaThrArgLYs 680  
DB 2070 TACATGTGACGCTTTCTGTTCACCTGAACATGATTTTGTGTGATGATCAACCGCAG 2129  
QY 681 GLYAenLYsIlleArgPheAlaAsnHISerValAsnProAsnCYeTYrAlaLYsVal 699  
DB 2130 GGTAACAATTCGTTTGCAATCATCTGTAATTCCAACTGCTATGCAAAAGTT 2186  
RESULT 6  
ADT90198  
ID ADT90198 standard; DNA; 2600 BP.  
XX  
AC ADT90198;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE Human gene over-expressed in a cancerous cell, SEQ ID 311.  
XX  
KM cancer; detection; pre-cancerous; anti-neoplastic; cancer therapy;  
XX  
KN cytostatic; gene therapy; gene; de.  
XX  
OS Homo sapiens.  
XX  
PN WO2004091548-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 15-APR-2004; 2004MO-US009289.  
XX  
PR 15-APR-2003; 2003US-0462895F.  
XX  
PA (AVAL-) AVALON PHARM INC.  
XX  
PI Strovel JW, Cain CB, Horrigan SK, Augustus M;  
XX  
DR WPI; 2004-784489/77.  
XX  
PT Diagnosing cancer in a mammal comprises determining amplification of a  
XX  
PT specific gene, e.g., serine/threonine kinase 6, in the genome of a  
XX  
PT mammal.  
XX  
PS Claim 3; SEQ ID NO 311; 92pp; English.  
XX  
CC The invention relates to a novel method for diagnosing cancer in a  
XX  
CC mammal. The method comprises determining amplification of specific genes,  
XX  
CC given in the specification, in the genome of a mammal. The invention  
XX  
CC further comprises methods for: diagnosing cancer or a pre-cancerous  
XX  
CC condition in a mammal; inhibiting cancer or a pre-cancerous condition in  
XX  
CC a mammalian cell; identifying an agent having therapeutic activity in a  
XX  
CC human patient; identifying an anti-neoplastic agent; determining the  
XX  
CC cancerous status of a cell; identifying a compound as an anti-neoplastic  
XX  
CC agent; treating cancer; monitoring the progress of cancer therapy in a  
XX  
CC patient; determining the likelihood of success of cancer therapy in a  
XX  
CC patient; producing test data with respect to the anti-neoplastic activity  
XX  
CC of a compound; and determining the progress of a treatment for cancer in  
XX  
CC a patient following commencement of a cancer treatment for the patient.  
XX  
CC The method and compositions of the invention have cytostatic activity  
XX  
CC and may be used in gene therapy. The method is useful in diagnosing or  
XX  
CC treating cancer in a mammal. This polynucleotide sequence represents a  
XX  
CC gene over-expressed in a cancerous cell, used in the novel cancer  
XX  
CC diagnosis method of the invention.  
XX  
SQ Sequence 2600 BP; 818 A; 525 C; 630 G; 627 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,386-282 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 13 Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x ADT90198 (1-2600)  
QY 1 MetGlyInThrGlyLYsLYsSerglYlYsGlyProValCYeSTRArgLYsArgLYs 20  
DB 90 ATGGCCAGACCTGGGAAGAAATCTGAGAGGAGCAAGTTTGTGCGGAGGCTGTAATA 149  
QY 21 SerglYlYsMetArgLeuArgGlnLeuLYsArgPheAlaGAlaAspGlnValLYsSer 40  
DB 150 TCAGAGTACATCCGACTGACAGCTCAAGAGGTTCAAGCAGCTGATGAAAGAGT 209  
QY 41 MetPheSerSerAsnArgGlnLYsIlleLeuGluArgThrGlnIleLeuAsnGlnGlnIuTrP 60  
DB 210 ATGTTAGTCCATGCTGACGAAATTTTGGAGAGACGAAATCTTAAACCAAGATCG 269  
QY 61 LYsGlnArgArgLYsGlnProValHISerLeuThrSerValSerSerLeuArgLYsThr 80  
DB 270 AAACAGCAGAGATACAGCTCTGTGACATCTGACTTCTGTAGCTCATTCGCGGAGCT 329  
QY 81 ArgGluCYsSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLYsThr 100  
DB 330 AGGAGGTTCGCTGACAGTCACTGATTTTCCAAACAAGTATCCCATTAAGACT 389  
QY 101 LeuAsnAlaValAlaSerValProIleMetTYrSerTrpSerProLeuGlnAsnPhe 120  
DB 390 CTGAATGAGATGCTTCAGTACCCATTAATGATTTTGTGCTCCCTTACAGCAGAAATTT 449  
QY 121 MetValGluAspGluThrValLeuHISerValProTYrMetGlyAspGlnValLeuAsp 140  
DB 450 ATGGTGAAGATGAACCTGTTTACATTAACCTCTTATATGGAGATGAAGTTTATGAT 509  
QY 141 GlnAspGlyThrPheIlleGlnGlnLeuIleLYsAsnTYrAspGlyLYsValHISerLYsAsp 160  
DB 510 CAGGATGCTACTTTCATTAAGAACTAATAAATTTATGATGGGAAAGTACACCGGAGT 569  
QY 161 ArgGluCYsGlyPheIlleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlnIuTrP 180  
DB 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTGTGTAATGCTTGTGCTCA 629  
QY 181 TYrAsnAspAspAspAspAspAspLYsAspAspProGlnGluArgGlnIuTrP 200  
DB 630 TATATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
QY 201 LYsAspLeuGluAspPheIleArgAspAspLYsGlnSerArgProProArgLYsPheProSer 220  
DB 690 AAAGATCTGAGAGATCAACGAGATGATTAAGAAACCCGACCTCGGAAATTTCTTCT 749  
QY 221 AspLYsIllePheGlnAlaIleSerSerMetPheProAspLYsGlyThrAlaGlnGlnLeu 240  
DB 750 GATTAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGAGCAGCAGGAAAGACTA 809  
QY 241 LYsGlnLYsTYrLYsGlnLeuThrGlnGlnGlnIleuProGlnLYsAlaLeuProProGlnLYs 260  
DB 810 AAGGAAATATTAAGAACTACACGAGACAGCTCCAGGCGACTTCTCTGAAATCT 869  
QY 261 ThrProAsnIleAspGlyProAsnAlaLYsSerValGlnArgGlnIleSerIleSer 280  
DB 870 ACCCCCAACATAGATGACCAATGCTAAATCTGTTGAGAGAGCAAGCTTACATCC 929  
QY 281 PheHISThrLeuPheCYeArgArgCYePheLYsTYrArgCYePheLeuHISProPheHIS 300  
DB 930 TTTCAATAGCTTTTCTGAGGAGATGTTTAAATGATGCTCTTCAATCCCTTTTAT 989  
QY 301 AlaThrProAsnThrTYrLYsArgLYsAsnThrGlnIleLeuAspAsnLYsProCYs 320

Db 990 GCAACACCCAACTTATTAACGGAGAACACAGAAACAGCTCTAGACAACTTGT 1049  
Qy 321 G1YProGInCySTyTgInH1sLeuGluG1Val1sYgU1PheAla1a1a1eUThrAla 340  
Db 1050 GGACCCAGAGTATACAGCATTTGGAGGAGCAAGAGTTTGCTGCTCTCAACGCT 1109  
Qy 341 G1uArG1leYThrPrOProLYaRgProG1Yg1ArGaRgArG1YArGLeuProAn 360  
Db 1110 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGAGCCGACAGAGAGCGCTCCAAAT 1169  
Qy 361 AAsnSerSerArGProSerThrProThr11eAsnVal1eUglUserLYaSPThrAspSer 380  
Db 1170 AACAGTACAGCCCAAGACCCCAACATTAATGCTGAGATCAAGAGATACAGACGT 1229  
Qy 381 AAspArG1uAlaG1YThrG1uThrG1Yg1Yg1uAsnAAsnAPLYg1UG1UG1uLYs 400  
Db 1230 GATAGGGAGAGCGGACTGTAAGCGGGGAGAGAAACATGATTAAGAGAGAGAGAG 1289  
Qy 401 LYsAPG1uThrSerSerSerSerG1uAlaAsnSerArGyG1nThrPro11eYMet 420  
Db 1290 AAAGTGAACCTTGAGCTCCTCTGAGCAAAATTCGCTGCAAAACCAATTAAGATG 1349  
Qy 421 LYsProAn11eG1uProProG1uAsnVal1G1uTpsSerG1YAlaG1uAlaSerMetPhe 440  
Db 1350 AAGCCAAATATTGAACCTCTGAGATGTGAGTGGAGTGGTGAAGCCTCAATGTTT 1409  
Qy 441 ArGVal1eU11eG1YThrTYrTYrAspAnPheCYsAla11eAlaArgLeu11eG1YThr 460  
Db 1410 AAGAGCTCATTTGGACCTTACTATACAAATTTCTTGCCATTGCTAGATTAAATGGACC 1469  
Qy 461 LYsThrCYsArG1nVal1YrG1uPheArgVal1Yg1uSerSer11e11a1aProAla 480  
Db 1470 AAAACATGTAGACAGGTGTATAGATTAGATCAAAAGATCTAGCATCATAGCTCAAGCT 1529  
Qy 481 ProAlaG1uAPrValAspThrProProArgLYsLYsLYsLYsLYsLYsLYsLYsLYsLYs 500  
Db 1530 CCCCCTGAGATGTGGATATCTCTCCAGAGAAAGAGAGAGAAACACCGGTGTGGGCT 1589  
Qy 501 AlAH1sCYsARGLYs11eG1uLeuLYsAPG1YSerSerAsnH1sVal1YrAsnTYr 520  
Db 1590 GCACCTGCAAGAAAGATACAGCTGAAAAAGAGCGCTCTCTTACCACTTTACACAT 1649  
Qy 521 G1nProCYsAPR1sProArG1nProCYsAPSerSerCYsProCYsVal11eAlaG1n 540  
Db 1650 CAACCTGTGATCATCCACGGCAGCTTGTCACATGTCGTGCCCTTGTTGTATAGACAA 1709  
Qy 541 AAsnPheCYsG1uLYsPheCYsG1nCYsSerSerG1uCYsG1nAAsnArG1PheProG1YCYs 560  
Db 1710 AATTTTGTGAAGATTGTTGTCMAATGTAGTTCAGAGTGCMAAACCGCTTTCCGGGATGC 1769  
Qy 561 ArGyCYsLYsAlaG1nCYsAsnThrLYsG1nCYsProCYsTYrLeuAlaValArG1uCYs 580  
Db 1770 CGCTGCAAGACAGTGCACACAGCAGAGTCCCGCTGCTACTGCTGCTCGAGAGTGT 1829  
Qy 581 AAspProAspLeuCYsLeuThrCYsG1YAla1a1aAspH1sTRPAspSerLYsAsnValSer 600  
Db 1830 GACCTGACCTGTGCTTACTTGTGTGAGCGCGTGCACATTTGGAGACATTAATAATGTGTCC 1889  
Qy 601 CYsLYsAsnCYsSer11eG1nArG1YSerLYsLYsLYsLYsLYsLYsLYsLYsLYsLYs 620  
Db 1890 TGCAAGAACTGCAATTCAGCGGGGCTCCAAAAAGCATTAATCTGCGACACATCTAGC 1949  
Qy 621 Val1aG1YrG1Y11ePhe11eLYsAPrOVal1G1nLYsAsnG1UPhe11eSerG1u 640  
Db 1950 GTGGAGGCTGGGGAGTTTATCAAAAGATCTGTGCGAGAAAAAGAAATTCATCTCAGAA 2009  
Qy 641 TYrCYsG1Yg1u11e11eSerG1nAspG1uAlaAspArGArG1YLYsVal1YrAspLYs 660  
Db 2010 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGAGGAAAGTGTATGATAAA 2069  
Qy 661 TYrMetCYsSerPheLeuPheAsnLeuAsnAsnAPrPheVal1ValAspAlaThrArgLYs 680  
Db 2070 TACATGTGAGCTTTCTGTTCAACTTGAACAATGATTTTGTGTGATGCAACCGGCAAG 2129

Qy 681 G1YAsnLYs11eArgPheAlaAsnH1sSerVal1aAsnProAnCYSTYrAlaLYsVal 699  
Db 2130 GGTAAACAAATTTGTTTGCAAATTCGTTAAATCCAACTGCTATGCAAAAGTT 2186  
RESULT 7  
ADQ22638  
ID ADQ22638 standard; DNA; 2711 BP.  
XX  
AC ADQ22638;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5458.  
XX  
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW de.  
XX  
OS Homo sapiens.  
XX  
PN WC0204048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PP 26-NOV-2003; 2003WC-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnick A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 5458; 210bp; English.  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cyostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2711 BP; 865 A; 547 C; 658 G; 641 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 2,52e-282 Length: 2711  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 12 Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x ADQ22638 (1-2711)  
Qy 1 MetG1YnThrG1YLYsLYsSerG1uLYsG1YProValCYSTPArG1YsArGValLYs 20  
Db 167 ATGGGCCAGACTGGAGAAATCTGAGAGAGGAGCAAGTTTGTGTGGCGAGACCGGTAAAA 226  
Qy 21 SerG1uTYrMetArGLeuArG1nLeuLYsArG1nLeuLYsArG1nLeuLYsArG1nLeuLYsSer 40

227 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTCCAGACGAGCTGATGAAAGATG 286  
QY MetPheSerSerAsnArgGlnIleuGluArgThrGluIleuAsnGlnGluTrp 60  
Db 41 MetPheSerSerAsnArgGlnIleuGluArgThrGluIleuAsnGlnGluTrp 60  
QY 287 ATGTTAAGTTCCAAATGTCAGAAATTTGGAAAGAACGGAATCTTAAACCAAGATG 346  
QY 61 LysGlnArgArgIleGlnProValHisIleuThrSerValSerSerLeuArgGlyThr 80  
Db 347 AAACCGGAGAGATACAGCTTGTGACATCTGACTCTGTGAGCTCATTTGGCGGAGACT 406  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuGlyThr 100  
Db 407 ACGAGAGTTGTCGTCAGAGTGAAGTTGATTTCCAAACAGTCAATCCATTAAAGACT 466  
QY 101 LeuAsnAlaValAlaSerValProIleMetIleTrpSerProLeuGlnIleAsnPhe 120  
Db 467 CTGAATGCAAGTTCCTGAGTACCCCAATGATCTTCGTGCTCCCTTACAGCAAAATTT 526  
QY 121 MetValGluAspGluThrValIleuHisAsnIleProIleMetGluAspGluValLeuAsp 140  
Db 527 ATGTCGAGAGATGAAATCTTTTACATACATCTCTTAATAGGAGATGAAGTTTAAAT 586  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleuLysAsnIleAspGlyLysValHisGlyAsp 160  
Db 587 CAGATGATGATCTTTGATGAAAGACTAATTAATAAATTTAGTGGGAAAGTTACACGGGAT 646  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 647 AGAGATGTCGGTTAATAATGATGAATTTTGTGGAGTGTGATGATCCCTTGGTCAA 706  
QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
Db 707 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 766  
QY 201 LysAspLeuGluAspHisAspAspAspGlyLysAspArgProProArgLysPheProSer 220  
Db 767 AAAGATCTGAGAGATCACCGAGATGATTAAGAAAGCCGCCACTCGGAATTTCTCTT 826  
QY 221 AspGlyIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
Db 827 GATTAATAATTTTGAAGCCATTTCTCAATGTTCCAGATAGAGGCAACAGCAAGAACTA 886  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnIleuProGluValAlaLeuProGluCys 260  
Db 887 AAGGAAATATATAAGAACTCACCGAACAGAGCTCCAGCGCACTTCTCTTGATAT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnIleuSerLeuHis 280  
Db 947 ACCCCCAATATGATGAGCAAAATGCTAAATCTGTTCAAGAGAGAGCAAACTTACACTC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisAspPheHis 300  
Db 1007 TTTCTACAGCTTTCTGTCGAGATGTTTAAATATGACTGCTCTCAATCCTTTTCAAT 1066  
QY 301 AlaThrProAsnThrTyrIleAspArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
Db 1067 GCAACACCCCAACACTTATAGCGGAGAAACACAGAAACAGCTCTAGACAAACAACTTGT 1126  
QY 321 GlyProGlnCysTyrGlnHisIleuGluGlnIleuValLysGluPheAlaAlaIleuThrAla 340  
Db 1127 GGAACCAAGCTGTTACAGATTTTGAAGGAGCAAAAGAGTTGCTGCTGCTTCAACCGCT 1186  
QY 341 GluArgIleLysThrProProLysArgProGlyLysArgArgArgGlyValArgLeuProAsn 360  
Db 1187 GAGCGGATTAAGACCCCAACAAACGTCACAGAGGCGCGAGAAAGAGAGCGCTTCCCAAT 1246  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValIleuGlnSerLysAspThrAspSer 380  
Db 1247 AACAGTAGAGGCGCCAGACCCCAACATTAATGTCGTAATCAAAGATACAGACAGT 1306  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluLys 400  
Db 1307 GATTAGGAAGCAAGGAGCTGAACCGGGGAGAGAAACATGATTAAGAAAGAAAGAGAAAG 1366

QY 401 LysAspGluThrSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
Db 1367 AAAGATGAACCTTCAGAGCTCCCTAGAGCAAAATTTCCGTGTCAACACCAATTAAGATG 1426  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
Db 1427 AAGCAATATATTGAACCTCTGAGAAATGTGAGATGAGATGAGTGTGAGACCTCAATGTTT 1486  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1487 AGAGTCTCATTTGGCACTTACTATGACAAATTTCTGTGCATTTGATGTTAATTTGGACC 1546  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
Db 1547 AAACATGTAGACAGGTGTAGATGATTAAGTCAAAATTCATGACATCAATCTCAGCT 1606  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisAspLeuThrAla 500  
Db 1607 CCGCGTAGAGATGTGATCTCTTCCAGAGAAAGAAAGAGAAACACCGGTTGTGGCT 1666  
QY 501 AlaHisCysArgLysIleGlnIleuLysAspGlySerSerAsnHisValTyrAsnTyr 520  
Db 1667 GCACACTGACAGAAATGATCAGCTGAAAGAGAGCGCTCCTTAACATGTTTACAACTAT 1726  
QY 521 GlnProCysAspHisAspArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1727 CAACCTGTGATCATCCAGGAGCCTTGTGACATGTCGTGCTTGTGTATACACAA 1786  
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
Db 1787 AATTTTGTGAAAGTTTGTTCATATGATGATGATGATGATGATGATGATGATGATGAT 1846  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
Db 1847 CGCTCAAGCAAGTGTGACACCAAGCAGAGTCCCGTCTACCTGCTGTCCGAGAGTGT 1906  
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAlaAspHisAspAspSerLysAsnValSer 600  
Db 1907 GACCTTGACCTTGTCTTACTGTGTGAGCCGCTGACATGTGGACAGTAAATATGTGTC 1966  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleuLeuAlaProSerAsp 620  
Db 1967 TGCAGAACTGCAAGATTCAGCGGGCTCCAAAAGCATATTTCTGGACCACTGAC 2026  
QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 2027 GTGGCAGGCTGGGAGATTTTATCAAGATCTGTGCAAGAAATGAATTCATCTCAGAA 2086  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
Db 2087 TACTGTGAGAGATTAATTTCTCAAGATGACCTGACAAAGAGGAAAGTGTATATAA 2146  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
Db 2147 TACATGTCAAGCTTTCTTTCACTTGAACATGATTTTGTGTGATGCAACCGCGAAG 2206  
QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
Db 2207 GGTAACAAAATTCGTTTGAATCAATTCGGTAAATTCAAACTGCTATGCAAAAGTT 2263

RESULT 8  
ADG10641  
ID ADG10641 standard; cDNA; 2558 BP.  
AC ADG10641;  
XX 26-FEB-2004 (first entry)  
DT Human STAT6-activating protein-encoding cDNA, SEQ ID NO:231.  
DE Human; STAT6-activating protein; drug screening; activator; inhibitor;  
KW allergic disease; inflammation; autoimmune disease; diabetes;



KM hyperlipidaemia; cancer; infection; HIV infection; antiallergic;  
KM human immunodeficiency; cancer; Th1 hyperfunction; antiallergic;  
KM antiinflammatory; antidiabetic; antipneumonic; antileukemic; anti-HIV;  
KM cytoskeletal; gene therapy; antisense therapy; ribozyme therapy; gene; ss.

OS Homo sapiens.

FN W020296943-A1.

PD 05-DEC-2002.

PF 22-MAY-2002; 2002WO-JP004949.

PR 25-MAY-2001; 2001JP-00157043.

PR 30-AUG-2001; 2001JP-00260681.

PR 10-OCT-2001; 2001JP-00313175.

XX (ASAH ) ASAH KASEI KOGYO KK.

PI Honda G, Matsuda A, Muramatsu S, Ishizawa K;

DR WPI: 2003-140442/13.

DR P-PSDB; ADG10642.

XX STAT6-activating proteins and encoded genes, applicable in diagnosis of

PT and developing drugs to treat allergic diseases, inflammations,

PT autoimmune diseases, diabetes, hyperlipidemia, infections e.g. HIV, and

PT cancer.

XX Claim 4; SEQ ID NO 231; 2080bp; Japanese.

CC The invention relates to 242 human STAT6-activating proteins and cDNAs  
CC encoding them (ADG10411-ADG10894) and to sequences with 95% or more  
CC homology to the STAT6-activating proteins and their encoding nucleic  
CC acids. The invention also relates to recombinant vectors and host cells  
CC comprising a STAT6-activating protein-encoding nucleic acid; the  
CC recombinant production of a STAT6-activating protein; an antibody  
CC specific for a STAT6-activating protein; antisense oligonucleotides and  
CC ribozymes targeted to nucleic acids encoding a STAT6-activating protein;  
CC methods of screening for activators or inhibitors of STAT6-activating  
CC proteins; drug compositions comprising a modulator of STAT6-activating  
CC protein activity or expression; and methods of treating patients by  
CC administration of the drug compositions. The STAT6-activating proteins,  
CC nucleic acids encoding them, and modulators of their activity or  
CC expression are useful in the diagnosis and treatment of allergic  
CC diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia,  
CC cancer, infections (e.g., HIV), cancer and disorders associated with Th1  
CC hyperfunction. The present sequence is related to the invention.

XX SQ Sequence 2558 BP; 808 A; 513 C; 612 G; 625 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,74e-282 Length: 2558  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 10 Indels: 0  
Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x ADG10641 (1-2558)

QY 1 MetGlyGlnThrGlyLysLeuSerGlyProValCysTrpArgLysArgValYys 20  
DB 58 ATGGGCGACGCTGGGAGAAATCTGAGAGGGAGCCAGTTGTTGGCGGAAAGGTGTAATA 117  
QY 21 SerGlyLysMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValYysSer 40  
DB 118 TCAGAGTACATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 177  
QY 41 MetPheSerSerPheAspArgLysLysIleLeuGluLysThrGluLysLeuAsnGlnGluTrp 60  
DB 178 ATGTTAGTTCATCGATCGTCAAGAAATTTTGAAGAGACGAAATCTTAACCAAGAAATG 237

QY 61 LysGlnArgArgGlnIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 238 AAACAGGAGAGAGATACAGCCCTGACACATCTGACCTTCTGACCTCAATGCGGGAGCT 297  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 298 AGGAGTGTTCGGAGACCACTGACTGATTTTCCACACAAAGTCACTTAATAAGCT 357  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTrpSerProLeuGlnAsnPhe 120  
DB 358 CTGAATGAGTGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 417  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProLysMetGlyAspGluValLeuAsp 140  
DB 418 ATGATGAGAAATGAACTGTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 477  
QY 141 GlnAspGlyThrPheIleGluGlnLeuIleLysAsnIleLysAspGlyLysValHisGlyAsp 160  
DB 478 CAGATGCTACTTTCATTTGAAAGAACTAATAAAATTAATGATGGGAAAGTACAGGGGAT 537  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 538 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGTGAATGCTTGTGTCA 597  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluValGluGluGln 200  
DB 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
QY 201 LysAspLeuGluAspPheIleArgAspAspLysGluSerArgProArgLysPheProSer 220  
DB 658 AAAGATCTGGAGATCACCGAGATGATTAATAAAGCCGCCCACTCGGAATTTCTTCT 717  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 718 GATTAATATTTTGGAGGCAATTTCTCAATGTTTCAATGATGAGGACACGAGAAAGACTA 777  
QY 241 LysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 260  
DB 778 AAGGAAAATATTAAGAACTACCGAACAGAGCTCCAGGCGGACTTCTCTGAAATGT 837  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnAspGluGlnSerLeuHisSer 280  
DB 838 ACCCCACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysLysArgCysPheLeuHisProPheHis 300  
DB 898 TTTGATATGCTTTTCTGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 957  
QY 301 AlaThrProAsnThrLysLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 958 GCACACACCAACATTAATAGCTTAAAGCAAGAACAGAGCTTAAAGCAACAACTTGT 1017  
QY 321 GlyProGlnCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 340  
DB 1018 GAGCACAGTGTATACAGCACTTTGGAGGAGCAAGAGATTTGCTGCTCTACAGCT 1077  
QY 341 GlnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 360  
DB 1078 GAGCGGATTAAGAACCCCAACAAAGCTCAGAGAGGCGCAGAGAGAGAGAGAGAGAGAG 1137  
QY 361 AsnSerSerArgProSerThrProThrLysAsnValLeuGluSerLysAspThrAspSer 380  
DB 1138 AACAGTACAGGCGCCAGACCCCACTTATATGCTGAGATCAAGAGATACAGACAGT 1197  
QY 381 AspArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 400  
DB 1198 CATAGGAG 1257  
QY 401 LysAspGluThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 420  
DB 1258 AAAGATGAATCTTGAAGCTCTCTGAAGCAATTTCTGATGAGCAACCAATTAAGATG 1317  
QY 421 LysProAsnIleGluProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440

```

DB 1318 AAGCGAATATGTAACCTCTGATGATGAGATGAGTGTCTCAAGCTCAATGTTT 1377
QY 441 ArgValLeuIleGlyThrTyrTrpAspPheCysAlaIleAlaArgLeuIleGlyThr 460
DB 1378 AAGGTCCTCATGGCACTTACTATGACAAATTTCTGTCATGCTGATTAATGGAGACC 1437
QY 461 LysThrCysArgGlnValTyrGluPheArgValIleGluSerSerIleIleIleAlaProIa 480
DB 1438 AAAACATGTAGACAGGTGTATGATTTAGTCAAGAACTAGCAATCATAGCTCAAGCT 1497
QY 481 ProAlaGluAspValAspThrProArgIleLysLysLysArgLysIleAspArgLeuTrpAla 500
DB 1498 CCGCGTAGAGATGTGATATCTCTCCAGAGAAAAGAAAGAGAAACCGCGTTGTGGCT 1557
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValIleTrpAsnTyr 520
DB 1558 GCACACTGCAGAAAGATACAGCTGAAAGAGACGGCTCCCTGACCATGTTTACACTAT 1617
QY 521 GlnProCysAspHisAspArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB 1618 CAACCTGTATCATCCAGCGACGCTGTGTGACAGTTCCGTGCTGTGTGTATGACACAA 1677
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560
DB 1678 AATTTTGTGAAAAGTTTGTCTCATGTATGATGATGATGATGATGATGATGATGATGATG 1737
QY 561 ArgCysLysAlaGlnCysAsnThrIleGlnCysProCysTyrIleuAlaValArgGluCys 580
DB 1738 CCTGTGAAGACACAGTGCACACACAGAGTGGCCCTGTACTGCTGTGCGAGAGTGT 1797
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisIleTrpAspSerLysAsnValSer 600
DB 1798 GACCTGACCTGTCTTACTGTGTGAGCCGCTGACCATTTGGACAGTAAATGTGTCC 1857
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysIleLeuLeuLeuAlaProSerAsp 620
DB 1858 TCCAGAAATGCAGATTCAGCGGGGCTCCAAAAGCACTATGCTGGCACCATGTGAC 1917
QY 621 ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGlu 640
DB 1918 GTGGCAGGCTGGGGGATTTTATCAAAAGATCCGTGTCAAGAAATGATTCATCTCAGAA 1977
QY 641 TyrCysGlyGlnIleIleSerGlnAspGluAlaAspArgArgGlyLysValIleTrpAspLys 660
DB 1978 TACTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGGGAAAGTGTATGATAAA 2037
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680
DB 2038 TACATGTGCAAGCTTCTGTTCATCTGAACATGATTTTGTGTGATGCAACCGCGAAG 2097
QY 681 GlysAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699
DB 2098 GGTAACAAATTCGTTTGCATATCATTCGTAAATCCAAACTGCTATCAGAAAGTT 2154

```

RESULT 9  
ABX76182  
ID ABX76182 standard; DNA; 2576 BP.

AC ABX76182;  
DT 02-APR-2003 (first entry)  
XX Lung cancer-associated polynucleotide #54.  
DE Lung cancer-associated polynucleotide #54.  
XX Lung cancer-associated polynucleotide; gene; ds; cytosstatic; emphysema;  
KW antiinflammatory; antiaesthetic; non-small cell lung cancer; atelectasis;  
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX Unidentified.  
OS  
XX

```

PN WO200286443-A2.
XX
XX 31-OCT-2002.
XX
XX 18-APR-2002; 2002WO-US012476.
XX
XX 18-APR-2001; 2001US-0284770P.
XX 10-MAY-2001; 2001US-0290492P.
XX 09-NOV-2001; 2001US-0339245P.
XX 13-NOV-2001; 2001US-0350666P.
XX 29-NOV-2001; 2001US-0334370P.
XX 12-APR-2002; 2002US-0372246P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Aziz N, Murray R;
XX WPI; 2003-093161/08.
XX P-PSDB; ABUS6461.
XX
XX Detecting a lung cancer-associated transcript in a cell from a patient
XX for treating lung cancer, by contacting a biological sample from the
XX patient with a polynucleotide that exhibits increased or decreased
XX expression in lung cancer.
XX
XX Claim 22; Page 231-232; 453pp; English.
XX
XX The invention relates to a method for detecting a lung cancer-associated
XX transcript in a cell from a patient, comprising contacting a biological
XX sample from the patient with a polynucleotide that selectively hybridizes
XX to a sequence that is at least 80 % identical to a gene that exhibits
XX increased or decreased expression in lung cancer samples. Lung cancer-
XX associated polynucleotides and polypeptides are used for identifying a
XX compound that modulates a lung cancer-associated polypeptide, for
XX inhibiting proliferation of a lung cancer-associated cell to treat lung
XX cancer in a patient and for treating a mammal having lung cancer by
XX administering a modulatory compound identified. The methods are useful
XX for treating lung cancer, such as small cell lung cancer, non-small cell
XX lung cancer or other benign or precancerous lesions, e.g. atelectasis,
XX emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
XX hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
XX bronchiectasis. The genes, polynucleotides and polypeptides are useful
XX for diagnostic purposes and as targets for screening for therapeutic
XX compounds that modulate lung cancer, such as antibodies. Sequences of the
XX ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
XX invention
XX
XX Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 6.8e-282 Length: 2576
XX Score: 3802.00 Matches: 698
XX Percent Similarity: 99.9% Conservative: 0
XX Best Local Similarity: 99.9% Mismatches: 1
XX Query Match: 99.8% Indels: 0
XX DB: 8 Gaps: 0
XX
XX US-10-773-302-2_COPY_1_699 (1-699) x ABX76182 (1-2576)
XX
XX 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20
XX 58 ATGGCGCAGACTGGAGAAATCTGAGAAAGGACCAAGTTTGTGGCGGAAAGCTGTAA 117
DB 118 TCAGAGTATCATGCGACTGACAGCTCAAGAGTTCAAGCGAGCTGATGAACTAAAGAGT 177
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValIleSer 40
DB 118 TCAGAGTATCATGCGACTGACAGCTCAAGAGTTCAAGCGAGCTGATGAACTAAAGAGT 177
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluAlaGlyThrGlnIleLeuAsnGlnLeuTrp 60
DB 178 ATGTTTAAATTCGAAACGTCAGAAATTTTGGAAAGAGAAATCTTAAACCAAGATGG 237
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80

```

Db 238 AACACGCGAAGATACAGCCTGTGCAATCTGACCTTGTGAGCTCATGGCGGAGACT 297  
Qy ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 298 AGGAGTGTGGTGACCACTGATCTTGATTTTCCACACAGATCATCCATTAAAGACT 357  
Qy LeuAsnAlaValIleSerValProIleMetYrsSerProLeuGlnGlnAsnPhe 120  
Db 358 CTGATGCACTGTGCTCAGTACCCCATATGTAATCTTGCTCCCTCACACAGAAATTTT 417  
Qy MetValGluAspGluThrValLeuHisAsnIleProYrMetGluAspGluValLeuAsp 140  
Db 418 ATGGTGAAGATGAAGAACTGTTTACATACATCTCTATATAGGAGATGAAGTTTACAT 477  
Qy GlnAspGlyThrPheIleGluGluLeuIleYrsAsnYrAspGlyLysValHisGlyAsp 160  
Db 478 CAGATGTACTTTCTATTGAAGAACTATTAATAATTTGATGGAAAGTACCGGGAT 537  
Qy ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 538 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGGTGAATGCCCTGTGCA 597  
Qy TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluGln 200  
Db 598 TATATATGATGATGACGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAAGACG 657  
Qy LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
Db 658 AAAGATCTGGAGATACCGAGATATTAAGAAACCCGCCACTCGGAATTTCTTCT 717  
Qy AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
Db 718 GATATAATTTTGGAGGCCATTTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 777  
Qy LysGluLysYrLysGluLeuThrGluGlnGluLeuProGluValAlaLeuProGluCys 260  
Db 241 AAGGAAAAATATTAAGAACTCACCGAACGAGCTCCAGCGCATCTTCTCTAAGT 837  
Qy ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
Db 838 ACCCCCAACATGATGACCAAAATCTTAATCTGTTCAGAGAGACAAAGCTTACATCC 897  
Qy PheHisThrLeuPheCysArgArgCysPheLysYrAspCysPheLeuHisProPheHis 300  
Db 898 TTTCTACAGCTTTCTGTAGCGATGTTTAAATGATGACTGCTCTCTCATCTTTTCAT 957  
Qy AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
Db 958 GCACACCCCAACACTTATTAAGCGAAGACACAGAAACGCTCTAGACACAAACCTTGT 1017  
Qy GlnProGlnCysYrGlnHisLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340  
Db 1018 GAGCACAGTGTTCACAGATTTGAGGAGGACAAAGAGTTGCTGCTCTCTCACCGCT 1077  
Qy GlnArgIleLysThrProProLysArgProGlyLysArgArgArgGlyValArgLeuProAsn 360  
Db 1078 GAGCGGATTAAGACCCCAACAAACGTCACGAGAGCGCGACGAGAGCGGCTTCCCAT 1137  
Qy AsnSerSerArgProSerThrProThrIleAsnValIleGluLysSerLysAspThrAspSer 380  
Db 1138 AACAGTTAGCGGCCACGACCCCAACATTAATGTGCTGGAATCAAGAGATACAGACGT 1197  
Qy AspAspGluAlaGlyThrGluThrGluGluLysAsnAspAspLysGluGluGluLys 400  
Db 1198 GATAGGAGACGAGGACTGAACGCGGAGAGAGAACATGATTAAGAGAGAGAGAGAG 1257  
Qy LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
Db 1258 AAAAGATGAACCTTCGAGCTCTCTGAAGCAAAATTCCTCGGTGTCAAACACCAATTAAGATG 1317  
Qy LysProAsnIleGluProGluLysValGluTyrPheSerGlyValGluAlaSerMetPhe 440  
Db 1318 AAGCAAAATATTAAGCTCTCTGAGAAATGAGATGAGTGTGTGAACCTCTCAATGTTT 1377

Qy 441 ArgValLeuIleGlyThrYrYrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1378 AGAGTCTCTATGTGCACTACTATGACAAATTCGTGTCCATYGTGATTAATTTGGAGCC 1437  
Qy LysThrCysArgGlnValIleThrGluPheArgValIleGluSerSerIleIleAlaProAla 480  
Db 1438 AAAACATGTAGACGAGTGTATGAGTTAGAGTCAAGAAATCTAGACATCATGCTCCAGCT 1497  
Qy ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTyrAla 500  
Db 481 CCCGCTAGAGATGTGATATCTCTCAAGGAAAAAGAAAGAAACACCGGTTGAGGCT 1557  
Qy AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnYr 520  
Db 501 GCACACTGCAAGAAAGATACAGCTGAAAAAGACCGCTCTTAACATGTTTACAACTAT 1617  
Qy GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 521 CAACCTGTGATCATCCACGCGACCTTGTGACAGTTCGTGCTGTGTGATAGCACAA 1677  
Qy AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
Db 541 AATTTTGTGAAAAAGTTTGTGCATGTATGTTCAAGTGTCAAAACCGCTTCCGGAGATGC 1737  
Qy ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrLysAlaValArgGluCys 580  
Db 1678 CGTGCAAAGCACAGTGCACACCAAGACAGTCCCGGTCTACCTGGCTGTCCAGAGATGT 1797  
Qy AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisThrAspSerLysAsnValSer 600  
Db 581 GACCTGACCTGTGTCTTCTGTGTGAGCCCTGACCATTTGGAGCAGTAAATAATGTCTCC 1857  
Qy CysLysAsnCysSerIleGlnArgLysLysLysLysLysLeuLeuAlaProSerAsp 620  
Db 1798 TGCAAGAACTGCAATTCACGCGGGGCTCCAAAAGACATCTATTGTGGCACCATCTGAC 1917  
Qy ValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 621 GTGGCAGGCTGGGGGATTTTATCAAGATCTCTGTGCAGAAATATGAAATTCATCTCAGAA 1977  
Qy TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValYrAspLys 660  
Db 1978 TACGTGAGAGATTAATTTCTCAAGATGACCTGACAGAGAGGAAAGTATGATTA 2037  
Qy TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgLys 680  
Db 2038 TACATGTGCACTTTCTGTTCACTTGAAACATGATTTTGTGTGATGCAACCGCAAG 2097  
Qy GlnLysAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysYrAlaLysVal 699  
Db 2098 GGTAAACAATAATCGTTTGGCAAAATCATTCGTTAAATCCAAATCTGTATGCAAAAGTT 2154

RESULT 10  
ADK66988  
ID ADK66988 standard; DNA; 2576 BP.  
AC ADK66988;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX  
XX Gene #78 for inhibitory RNA to manipulate stem cell phenotype.  
XX de; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
XX pluripotent stem cell.  
XX Homo sapiens.  
XX OS  
XX PN WO2003068961-A2.  
XX 21-AUG-2003.  
XX PD 12-FEB-2003; 2003WO-GB000579.  
PF



Db 1678 AATTTTGTGAAAAGTTTGTCAATGTAGTTCAGAGTGTCAAAACCGCTTCCGGAGTGC 1737  
Qy 561 ArgCysIysAlaIagInCysAanThrlYsgInCysProCysTyrLeuAlaValArgGluCys 580  
Db 1738 CGCTGCAAGACAGTGTGACACCAAGAGTGCCTCGTCTACCTGGCTGTCCGAGAGTGT 1737  
Qy 581 AspProAapLeuApsLeuThrlCysGlyAlaIaAapHisTyrPaspSerIysAanValSer 600  
Db 1798 GACCTGACCTCTGTCTACTTGTGAGCCCGCTGACCATTTGGACAGTAAAAATGTGTCC 1857  
Qy 601 CysIysAanCysSerIleGlnArgGlySerIysIysHisLeuLeuLeuAlaProSerAsp 620  
Db 1858 TGCAGAAAGTCAAGATTCACGGGGCTCCAAAAGCATCTCATTTGTGGCACCACTGTGAC 1917  
Qy 621 ValAlaGlyTTPGlyIlePheIleIysAapProValGlnIysAanGluPheIleSerGlu 640  
Db 1918 GTGGAGGCTGGGGATTTTATCAAAAGATCTGTGCAAAAATGAAATTCATCTCAGAA 1977  
Qy 641 TyrCysGlyGluIleIleSerGlnAapGluIaAapArgArgGlyIysValTyrAapIys 660  
Db 1978 TACTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGGGGAAAGTGATGATGATAA 2037  
Qy 661 TyrMetCysSerPheLeuPheAanLeuAanAanAapPheValIaAapAlaThrArgIys 680  
Db 2038 TACATGTGCACCTTCTCTTCAACTGAAACAATGATTTTGTGTGGATGCAACCGCAG 2097  
Qy 681 GlyAanIysIleArgPheAlaAanHisSerValAanProAanCysTyrAlaIysVal 699  
Db 2098 GGTAACAATAATTCGTTTGCANAATCATTCGGTAATCCAAACTGCTATGCAAAAGTT 2154

RESULT 11  
ADN39391  
ID ADN39391 standard; cDNA; 2576 BP.  
AC ADN39391;  
XX 17-JUN-2004 (first entry)  
DT  
XX  
DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:875.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;  
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KW inflammatory disease; autoimmune disease;  
KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
KW detection; diagnosis; prognosis; drug screening; drug targeting;  
KW wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
KW vulnery; gene therapy; vaccine; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO2003042661-A2.  
XX  
PD 22-MAY-2003.  
XX  
XX 13-NOV-2002; 2002WO-US036810.  
XX  
XX 13-NOV-2001; 2001US-0350666P.  
XX 21-NOV-2001; 2001US-0334664P.  
XX 29-NOV-2001; 2001US-0334393P.  
XX 03-DEC-2001; 2001US-0335394P.  
XX 14-DEC-2001; 2001US-0340376P.  
XX 08-JAN-2002; 2002US-0347211P.  
XX 10-JAN-2002; 2002US-0347349P.  
XX 08-FEB-2002; 2002US-0355250P.  
XX 13-FEB-2002; 2002US-0356714P.  
XX 20-FEB-2002; 2002US-0359077P.  
XX 29-MAR-2002; 2002US-0368809P.  
XX 04-APR-2002; 2002US-0370110P.  
XX 12-APR-2002; 2002US-0372246P.  
XX 05-JUN-2002; 2002US-0386614P.  
XX 16-JUL-2002; 2002US-0396839P.  
XX 22-JUL-2002; 2002US-0397757P.

PR 22-JUL-2002; 2002US-0397845P.  
PR 09-SEP-2002; 2002US-0409450P.  
XX  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
PA After D, Aziz N, Gineburg NM, Gish KC, Glynn R, Hevezl PA;  
PI Mack DH, Murray R, Watson SB, Wilson KE, Zlocznik A;  
XX  
XX WPI; 2003-468649/44.  
DR P-PSDB; ADN39392.  
XX  
PT Determining the presence or absence of a pathological cell in a patient,  
PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
PT a nucleic acid in a biological sample.  
XX  
XX Claim 8; SEQ ID NO 875; 1385bp; English.  
XX  
CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularisation syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a nucleic acid sequence of the invention.  
XX  
SQ Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 6.8e-282 Length: 2576  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.8% Indels: 0  
DB: Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x ADN39391 (1-2576)  
Qy 1 MetGlyGlnThrGlyIysIysSerGluIysGlyProValCysTyrArgIysArgValIys 20  
Db 58 ATGGGCCAGACTGGAGAGAAATCTGAGAAAGGACCACTTTTGTGGCGGAACCGTGTAAA 117  
Qy 21 SerGluTyrMetArgLeuArgGlnLeuIysArgPheArgAlaAapGluValIysSer 40  
Db 118 TCAGAGTACATGACGACGTGACAGCTCAAGAGGTTCAAGACGATGAAGTAAGTAAAGT 177  
Qy 41 MetPheSerSerAanArgGlnIysIleLeuGlnIysArgIysIleAanGlnGluThr 60  
Db 178 ATGTTAGTTCCATTCGTCAGAAAATTTGAAAAGAACGAAATCTTAACCAAGAAATGG 237  
Qy 61 LysGlnArgArgGlnIleGlnProValHisIleLeuThrSerValIserSerIysArgGlyThr 80  
Db 238 AAACAGCAAGAGATACAGCTGTGACATCTCTGACTCTGAGACTCATTCGCGGGAGACT 297  
Qy 81 ArgGluCysSerValThrSerAapLeuAapPheProThrGlnValIleProLeuIysThr 100  
Db 298 AGGAGGTTCGGTGACAGAGTGAATTTTCAACACAAGATCATCCATTAAGACT 357  
Qy 101 LeuAanIaValIaIaSerValProIleMetTyrSerTyrPaspProLeuGlnIaAanPhe 120  
Db 358 CTGAATGCAATGCTTCAGTACCAATATGATTTCTGTGTCCTCCCTACAGCAATTTT 417  
Qy 121 MetValGluAapGluThrValLeuHisAanIleProTyrMetGlyAapGluValLeuAap 140

Db	418	ATGGTGGAAAGATGAACGTTTAACTAAACATTCCTTAAATGGAGATGAAGTTTAACT	477
QY	141	GLNAPGGLYTHPHEII EGIUGLIEULIELYSANTRYAPGLYLSVALHISGLYAP	160
Db	478	CAGGATGGATCTTTCATTTGAAAGACATAATAAAAATATGATGGAGAAAGTACACGGGAT	537
QY	161	ATGGIUCYSGLI PHEII EASNAPGLIIEPHEVALGILUEVALASNAI ALEULIYGLN	180
Db	538	AGAGAAATGGGTTTATATATATGATAAATTTTGAGAGTTGGTAAATGCCCTTGCTCA	597
QY	181	TYRAENASPAASPAASPAASPAAPGLIYAAPAPPCOGIUGLI AARGIUGLIUYSGLN	200
Db	598	TATATATGATGATGACGATGATGATGAGACGATCTCTGAAGAAAGAGAGAAAGACAG	657
QY	201	LYSAPAPLEGLI AAPHII EARGAPAPLYSGIUSERATGPROCTARGLYSPHEPROSER	220
Db	658	AAAAGATCTGGAGAGATCACCGAATGATATAAGAAACCCGCCCACTCGGAAATTTCTCT	717
QY	221	ASPLYSIIEPHEGLI AII ESETERMET PHEPROAPLYISGLIYTHRIAGIUGLIUEU	240
Db	718	GATAAAAATTTGGAGGCCATTTCTCCATATGTTTCAGATTAAGGGACAGCAAGAAACTA	777
QY	241	LYSGIULSYTYRGLIUEUTHRGILUGINGLIUEUPROGLIYALAEUPROGLIUCYS	260
Db	778	AAGGAAAAATATTAAGAACTACCGCAACGACGCTCCAGGCCCACTTCTCTGAAATG	837
QY	261	THRPROASNI EASPGIYPROASNAI AYSSEVALGIDATRGILINSEIUEHIISEER	280
Db	838	ACCCCAACATATGATGAGCAAAATGCTAATCTGTTCAGAGAGCAAAAGCTTACATCC	897
QY	281	PHEHISITHLEUPHECYAARGARGCYSPHELYSTYAAPCYSPHELEHIEPROBHEHIS	300
Db	898	TTTCATACGCTTTCTCTGTAAGCGAATGTTTAAATATGATGCTGCTTCTTCAATCCTTTCAT	957
QY	301	ALATRRPROASNTHRYLYLSARGLYSANTHRGILUTRAI ALEUASPAASLYSPROCY	320
Db	958	GCAACACCCACACTTAATTAAGCGAGAAACACAGAAACAGCTCTTGACACAAACCTTGT	1017
QY	321	GLYPROGINCYSTYRGLINHISLEUGLIUGLIYALYSGIUPHEAII AII AII ALEUTHRI A	340
Db	1018	GGACCAACGTTGTTTACACAGATTTGGAGGGAGCAAGATGTTTGGCTGCTCTCACGCT	1077
QY	341	GLIATGII ELYSTHTRPROPROLYSARGPROGLIYATGATGARGGLIYARGLEUPROASN	360
Db	1078	GAGCGGAATAAAGACCCCAACAAACGTCACAGAGGCCCAAGAGAGACGCGCTTCCCAAT	1137
QY	361	ASNSERSERRARGPROSETHTRPROTHRILEASNVALIEUGIUSERLYSAPSTHASPSE	380
Db	1138	AACAGTACACAGGCCCAAGACCCCAACATTAATGCTGCGAATCAAAAGATATACAGACGT	1197
QY	381	ASPARGGLI AII AGLIYTHRGILUTHRGIIYGLIYGLIUEASNAAPLYSGIUGIUGLIUYS	400
Db	1198	GATAGGAGACAGGACCTGAACCGGGGGAGAGAACATGATTAAGAAAGAAAGACAGAG	1257
QY	401	LYSAPGILUTHRSERSESSERSESGIULI AANSEARARGYSGIINTHPROLI ELYSNET	420
Db	1258	AAAAGATGAACCTTGACGCTCTCTGAAACCAATTTCTCGGTGCAAAACCAATTAAGATG	1317
QY	421	LYSPROASNIIEGLIUPROPROGLIUEASNVALGIUTPSESGIYALAGLI AII ESEMECPHE	440
Db	1318	AAGCCAAATATTTGAACCTCTCGAAGATGTGAGAGAGATGTGTCTGAAGCCTCAATGTTT	1377
QY	441	ARGVALIEULIEGLIYTHRYTYRASAPAPHECYSAI AII AII AARGLEULIEGLIYTH	460
Db	1378	AGAGTCTCATTTGGCACTTACTATGACAAITTCGTGCCATTTGCTAGGTTAATGGGACC	1437
QY	461	LYSTHRCYBARGLINVALIYRGLIUEHARGVALIYSGIUSERSEIIEII AII AII APROLA	480
Db	1438	AAAACATGTACACAGGTGATGAGATTTTGAATCAAAAGATCTAGCATCATATGCTCCAGCT	1497
QY	481	PROAIIAGIUAAPVALASPSTHTRPROCTARGLYSGLYLSVALGII AARGLEULITRAI A	500
Db	1498	CCCGCTGAGATGATGGATATCTCTCCAGAGAAAAAGAGAGAAACACCGGTTTGCGCT	1557

QY	501	AlAhIaGcYsAsgLyIleGInLeuIyLeYsApRgIySsErSaRnHIsVaIlyrASnTYr	520
Db	1558	GCACtGTCAGAAAGATRCACtGTGAAAAGACCGGCTCTTAAcCATGTTTACATAT	1611
QY	521	GInPrOCysApRHIsPPrOArGInPrOCysApSsErSerCySPrOCysValIleAIn	540
Db	1618	CAACCTGTGATTCATCCACGGCAGCTTGTCAGAGTTGTGTCCTCTGTGTATGACAA	1677
QY	541	AsnPhcYsGInUlyrPhcYsGInCySsErSerGluCySgInAsnAlyrPhePrOgLyS	560
Db	1678	AAATTTCGTGAAAAGTTTGTCAATGTAGTTCAAGTGTCAAAACCGCTTTCGGGATGC	1737
QY	561	ArgCySlySAlaGInCySAsnThlyrSgInCySPrOCysTyrlEuaIaValArGInUlyS	580
Db	1738	CGCTGCAGAACAGACGTGCACACCAAGCAGTGCCTGCTACCTGCTGTCCAGAGTGT	1797
QY	581	AsPPrOAsPLeUCysleUthrCySgIyAlaIaAsPHIsrTPAsPserLyAsnValSer	600
Db	1798	GACCTGACCTCTGTCTTACTTGTGTAGAGCGCTGACCATTTGGACAGTAAAAATGTGCC	1857
QY	601	CysIySAsnCySsErIleGInArGlySerIySlySHIsleUleUleuAlaPrOsErAsP	620
Db	1858	TGCAGAACTGCAGATTCAGCGGGGGCTCCAAAAGCATCTATGCTGGCACATCTGAC	1917
QY	621	ValIaGlyTPGlyIlePheIleLyAsAPPrOValGInLySAsnGInPheIleSergIu	640
Db	1918	GTGGAGGGCTGGGGATTTTATCATAGATCTGTGCAGAAAATGAATTCATCTCAGAA	1977
QY	641	TyrCySgIyGInUleIleSergInAsPgluaIaAsPArGArGlyLySValTyTrAsPlyS	660
Db	1978	TACTGTGGAGAGATTAATTTCTCAAGATGAGCTGACAGAAAGGGAAAGTATGATAA	2037
QY	661	TyrMetCySsErPheUlePheAsnLeuAsnAsnAsPheValIaAsPAlaThrArGlyS	680
Db	2038	TACATGTGCAGCTTTCGTGTCACTTGAAcCATGATTTTGTGTGTGATGCACCCGCAAG	2097
QY	681	GIyASnTySleArGPheAlaAsnHIsSerValAsnPrObAcCySTyrlAlaLySVal	699
Db	2098	GGTAAcCAAAATTCGTTTTCGAAATCATTCGGTAAATCCAAACTCTATGCAAAAGTT	2154
RESULT 12			
ADR25182			
ID	ADR25182	standard; DNA; 2576 BP.	
AC	ADR25182;		
DT	21-OCT-2004	(first entry)	
XX		Breast cancer prognosis marker #1043.	
XX			
KM	ds; breast cancer; prognosis; gene expression; diagnosis.		
XX			
XX	Homo sapiens.		
OS			
PN	WO2004065545-A2.		
XX			
XX	05-AUG-2004.		
PD			
XX	15-JAN-2004; 2004MO-US001100.		
PF			
XX	15-JAN-2003; 2003US-00342887.		
PR			
XX	(ROSE-) ROSETTA INPHARMATICS LLC.		
PA	(NECA-) NETHERLANDS CANCER INST.		
XX			
P1	Van't Veer LJ, He Y;		
XX			
DR	WPI; 2004-593473/57.		
XX			
PT	Classifying a breast cancer patient according to prognosis comprises determining the similarity between the level of expression of each of five genes in a cell sample taken from patient, to control levels.		
PT			

Classifying a breast cancer patient according to prognosis comprises determining the similarity between the level of expression of each of five genes in a cell sample taken from patient, to control levels.

XX	PS	Disclosure; SEQ ID NO 1043; 226bp; English.			
XX	CC	The invention relates to a method of classifying a breast cancer patient according to prognosis by determining the similarity between the level of expression of each of five genes for which markers are listed in the specification, in a cell sample taken from the breast cancer patient, to control levels of expression for each respective five genes to obtain a patient similarity value. The methods are useful for classifying a breast cancer patient according to prognosis. Kits and computer program products are useful for data analysis using the diagnostic, prognostic and statistical methods of the invention. This sequence corresponds to a marker used in the method of the invention.			
XX	CC	Sequence 2576 BP; 826 A; 513 C; 612 G; 625 T; 0 U; 0 Other;			
	Alignment Scores:				
	Pred. No.:	6.8e-282	Length:	2576	
	Score:	3802.00	Matches:	698	
	Percent Similarity:	99.9%	Conservative:	0	
	Best Local Similarity:	99.9%	Mismatches:	1	
	Query Match:	99.8%	Indels:	0	
	DB:	13	Gaps:	0	
US-10-773-302-2_COPY_1_699 (1-699) x ADR25182 (1-2576)					
QY	1	MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys	20		
DB	58	ATGGCCGACACTGGAGAAATCTGAAAGGACCAAGTTGTGGCGAAGCGGTATAA	117		
QY	21	SerGlnYrMetArgLysArgLysGlnLysArgPheArgValLysArgLysSer	40		
DB	118	TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTCAGACGAGCTGATGAAGTAAAGGT	177		
QY	41	MetPheSerSerAsnArgGlnLysLysLysGlnValGlnGlnLysGlnLysGln	60		
DB	178	ATGTTTACTTCCAATCGTCAGAAATTTGAAAGAAAGGAAATCTTAAACCAAGATGG	237		
QY	61	LysGlnArgArgLysGlnProValHisLysLeuThrSerValSerSerLysArgLys	80		
DB	238	AAACGCGAAGATACAGCTCTGTCATCTTGAGTCTGAGTCTGAGTCTGAGTCT	297		
QY	81	ArgGlnCysSerValThrSerAspLysAspPheProThrGlnValLysProLys	100		
DB	298	AGGAGGTGTCGGTACCACTGATCTTCCATCAACCAAGTATCCCATTAAGACT	357		
QY	101	LeuAsnAlaValAlaSerValProLysMetYrSerTrpSerProLysGlnLysPhe	120		
DB	358	CTGAATGAGTGTCTTCAGTACCCAAATGTAATCTTGGTCTCCCTACAGCAGAAATTT	417		
QY	121	MetValGluAspGluThrValLeuHisAsnLysProYrMetGluAspGluValLys	140		
DB	418	ATGCTGGAGATGAATCTGTTTACATTAACATCTCTTATATGGAGATGAAGTTTAAAT	477		
QY	141	GlnAspGlyThrPheLysLysGlnLysLysAsnYrAspGlyLysValHisGlyAsp	160		
DB	478	CAGATGTGATCTTCAATGAAGAACTTAATAAATTAATGATGGAAAGTACACGGGAT	537		
QY	161	ArgGlnCysGlyPheLysAsnAspGluLysPheValGlnLysValLysAlaLysGln	180		
DB	538	AGAGATGTGGGTATTAATGATGAATTTTGTGGAGTGGTAAATGCTTGGTCAA	597		
QY	181	TyrAsnAspAspAspAspAspAspGluAspAspProGlnGluLysGln	200		
DB	558	TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	657		
QY	201	LysAspLysGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer	220		
DB	658	AAAGATCTGAGAGATCAACGAGATGATAAAGAAACCGCCACCTCGGAAATTTCTCT	717		
QY	221	AspLysLysPheGlnAlaLysSerMetPheProAspLysGlyThrAlaGlnGlnLys	240		
DB	718	GATTAATTTTGGAGGCACTTCTCAATGTTTCCAGATTAAGGCAACGAGAAAGAACTA	777		
QY	241	LysGlnLysYrLysGlnLysGlnGlnLysGlnLysGlnLysGlnLysGlnLysGln	260		
DB	778	AAAGAAATTAATTAAGAACTACCGAAGACAGAGTCTCCAGCGCATCTCTCGAATGT	837		
QY	261	ThrProAsnLysAspGlyProAsnAlaLysSerValGlnArgGlnLysSerLysSer	280		
DB	838	ACCCCAACATGATGAGCAAAATGCTTAATCTGTTCAGAGAGCAAGCTTACATCTCC	897		
QY	281	PheHisThrLysPhePheCysValArgCysPheLysYrAspCysPheLysHis	300		
DB	898	TTTCAATGCTTTTCTGTAGGCGATGTTTAAATATGATGCTTCTCAATCTTTTCAAT	957		
QY	301	AlaThrProAsnThrYrLysArgLysAsnThrGlnThrAlaLysAsnLysProCys	320		
DB	958	GCAACCAACCAACTTATTAAGCGGAGACAGAAACAGCTTACCAACCACTTGT	1017		
QY	321	GlyProGlnCysYrGlnHisLysGlnGlyValLysGlnLysPheAlaLysThrAla	340		
DB	1018	GGACCAAGTGTATCCAGCATTTGGAGGAGCAAGAGATTTGCTGCTCTTCAACGCT	1077		
QY	341	GlnArgLysLysThrProProLysArgProGlyGlyYrArgArgArgGlyArgLys	360		
DB	1078	GAGCGATTAAGACCCCAACCAACCTCCAGAGAGCCGCAAGAGAGAGCGCTTCCCAT	1137		
QY	361	AsnSerSerArgProSerThrProThrLysAsnValLysGlnLysLysPheThrAspSer	380		
DB	1138	AAACATGACAGCCCAACCAACCTTCAATATGCTGAGATCAACCAATTAAGATG	1197		
QY	381	AspArgGlnLysGlnLysThrGlnGlyGlyLysAsnAspLysGlnGlnLysGlnLys	400		
DB	1198	GATGGAGAGCAAGGACTGAAACCGGGAGAGAACTATTAAGAAAGAAAGAGAG	1257		
QY	401	LysAspGlnThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProLysMet	420		
DB	1258	AAAGATGAATCTTGAGCTCTTGAAGCAAAATCTCGGTGTCAACCAACCAATTAAGATG	1317		
QY	421	LysProAsnLysGlnProProGlnAsnValGlnTrpSerGlyAlaGlnLysMetPhe	440		
DB	1318	AAAGCAATATTAATTAATCTTCAAGAAATGAGAGAGTGTGTGAAGCTTCAATGTTT	1377		
QY	441	ArgValLysLysGlnThrYrYrAspAsnPheCysAlaLysLysLysLysLysLys	460		
DB	1378	AGAGTCTCATGAGCACTTACTATGACAATTTCTGTGCAATCTGCTAGTTAATGGAGC	1437		
QY	461	LysThrCysArgGlnValYrGlnPheArgValLysGlnSerSerLysLysLysLys	480		
DB	1438	AAACATGTAACAGAGTGTATGATTAAGTCAAAAGATCTACATCATGCTCCAGCT	1497		
QY	481	ProAlaGluAspValAspThrProProArgLysLysLysLysLysLysLysLysLys	500		
DB	1498	CCCCCTGAGATGAGTACTCTTCCCAAGAAAGAAAGAAAGAAACCGGTTGTGGCT	1557		
QY	501	AlaHisCysArgLysLysGlnLysLysAspGlyLysSerSerAsnHisValYrAsnYr	520		
DB	1558	GCACTGCAAGAAAGATACAGCTAATAAAGAGAGAGCTCTTAAACATGTTTACATAT	1617		
QY	521	GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValLysAlaGln	540		
DB	1618	CAACCTGTATCATCCAGCGCAAGCTTGTGACAGTTGCTGCTGCTGTGTATGACAA	1677		
QY	541	AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnAspPheProGlyCys	560		
DB	1678	AATTTTGTGAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGAT	1737		
QY	561	ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrLysAlaValArgGlnCys	580		
DB	1738	CGCTGCAAGCAAGTGAACCAAGAGAGTGGCCGCTCTCTGCTGCTGCTGCTGCTGCT	1797		
QY	581	AspProAspLysCysLysThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer	600		
DB	1798	GACCTGACCTCTGTCTTACTTGTGAGCGGCTGACCATTTGGGACAGTAAATATGTGCTC	1857		





QY 281 Pheh1sthrleuPheCySarGArCySphelyfYzAspCySphelenuh1sProPheh1s 300  
Db 898 TTTTCATACGCTTTCTCTGTAGCGATGTTTTAAATATAGCTGCTTCATCATCTTTTCAT 957  
QY 301 A1AthrProAnthrTYrLYsArGLyAsnThrG1uThrAlaLeuAspAsnLYsProCyS 320  
Db 958 GCACACCCCAACACTTATAGCGAGAGAACACAGAAACAGCTCTAGACACAAACCTTGT 1017  
QY 321 G1YProG1nCySfYrG1nH1sLeuG1uG1yAlaLYsG1uPheAlaAlaLeuThrAla 340  
Db 1018 GGACACAGAGTGTACAGCACTTGTAGAGAGACAAAGAGTTGCTGCTCTACCGCT 1077  
QY 341 G1uArG1leLYsThrProProLYsArProG1yG1yArGArGArG1yArG1eUProAsn 360  
Db 1078 GAGCGCATTAAGACCCCAACAAAGCTCCAGAGAGCGCGAGAGAGAGCGGCTTCCCAT 1137  
QY 361 AsnSerSerArGProSerThrProThr1LeAsnValLeuG1uSerLYsAspThrAspSer 380  
Db 1138 AACAGTAGCAGGCCACAGACCCCAACATATGTGTGAGATCAAGATACAGACAGT 1197  
QY 381 AspArG1uAlaG1yThrG1uThrG1yG1yG1uAsnAspLYsG1uG1uG1uLYs 400  
Db 1198 GATGGAGAGAGGAGGAGTGAACGGGGGAGAGAACATATGAATTAAGAGAGAGAGAG 1257  
QY 401 LYsAspG1uThrSerSerSerSerG1uAlaAsnSerArGArG1nThrPro1LeLYsMet 420  
Db 1258 AAAGTGAACCTTGAGCTCTCTGAGACAAATTCGTGGTCTCAACACCAATTAAGATG 1317  
QY 421 LYsProAsn1LeG1uProProG1uAsnValG1uTPSerG1yAlaG1uAlaSerMetPhe 440  
Db 1318 AAGCCAATATATTGAACCTCTGAGAGATGTGAGTGTGGTGTGAGAGCTCAATGTTT 1377  
QY 441 ArGValleu1LeG1yThrTYrAspAsnPheCySAla1LeAlaArGLeu1LeG1YThr 460  
Db 1378 AGAGCTTCATTTGGACTTACTATGACAAATTCGTGCAATTCCTAGGTATTTGGACC 1437  
QY 461 LYsThrCySArG1nValTYrG1uPheArGValLYsG1uSerSer1Le1AlaProAla 480  
Db 1438 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAGCATCATAGCTTCAGCT 1497  
QY 481 ProAlaG1uAspValAspThrProProArG1yLYsLYsLYsLYsLYsLYsLYsLYsLYs 500  
Db 1498 CCCCCTGAGAGTGTGATATCTCTCCAGAGAAAGAGAGAGAGACACGGGTGGGGCT 1557  
QY 501 AlaH1sCySArG1yS1LeG1uLeuLYsAspG1ySerSerAsnH1sValTYrAsnTYr 520  
Db 1558 GCACCTGACAGAAAGATACGTGAAAAGAGCGGCTCTTACCATGTTTACACTAT 1617  
QY 521 G1nProCySAspRn1sProArG1nProCySAspSerSerCySProCySVal1LeAlaG1n 540  
Db 1618 CAACCTGTGATCATCCAGCGAGCTTGTGACATGTTGTGCTGTGATAGACAA 1677  
QY 541 AsnPheCySGLYsPheCySGLYsSerSerG1uCYrG1nAsnArPheProG1YCyS 560  
Db 1678 AATTTTGTGAAGATTTTGTCAATGTATGTCAAGTGTCAAAACCGCTTCCGGGAGTC 1737  
QY 561 ArG1yLYsAlaG1nCYsAsnThrLYsG1nCYsProCYrTYrLeuAlaValArG1uCYs 580  
Db 1738 CGCTCAAGACACAGTGTGAACCAAGAGTCCCGTGTACTCTGCTGCTCGAAGTGT 1797  
QY 581 AspProAspLYsCyLeuThrCYsG1yAlaAlaAspRn1sTRPAspSerLYsAsnValSer 600  
Db 1798 GACCTTGACCTTGTCTTACTGTGTGAGCGCTGTACATTTGGAGACAGTAAATGTGTCC 1857  
QY 601 CySLYsAsnCYsSer1LeG1nArG1ySerLYsH1sLeuLeuLeuAlaProSerAsp 620  
Db 1858 TGCAAGAACTGCAAGTATTCAGGGGAGCTCAAAAGCATCTATTCTGGCACCATCTGAC 1917  
QY 621 ValAlaG1YrTG1Yr1LePhe1LeLYsAspProValG1nLYsAsnG1uPhe1LeSerG1u 640  
Db 1918 GTGGCAGGCTGGGGATTTTATCAAAAGATCTGTGCGAGAAATGAATTCATCTCAGAA 1977

QY 641 TyrCyG61yG1u1Le1LeSerG1nAspG1uAlaAspArGArG1yLYsValTYrAspLYs 660  
Db 1978 TACTGTGAGAGATTTTCTCAAGATGAGCTGACAGAGAGAGAGAGATGATGAATA 2037  
QY 661 TyrMetCySserPheleuPheAsnLeuAsnAspPheValValAspAlaTRArGLyS 680  
Db 2038 TACATGTGACGCTTCTCTGTTCAACTTGACAAATTTTGTGTGATGTCACCCGCAAG 2097  
QY 681 G1yAsnLYs1LeArGphea1AsnH1sSerValAsnProAsnCYrTYrAlaLYsVal 699  
Db 2098 GGTACAAATTCGTTTGTCAATTCATTCGTTAAATCCAAATCGTATGCAAAAGTT 2154  
RESULT 14  
ADP95921  
ID ADP95921 standard; DNA; 2253 BP.  
XX  
AC ADP95921;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human E2H2 (enhancer of zeste homologue 2)-encoding DNA, SEQ ID NO:2.  
XX  
KW Human; E2H2; enhancer of zeste homologue 2; chromosome 7q35;  
KW histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
KW inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
KW lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
KW prophylaxis; prevention; diagnosis; cytostatic; gene; de.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 1..2253  
FT CDS /\*tag= a  
FT /partial  
FT /product= "E2H2 (enhancer of zeste homologue 2)"  
FT /note= "No stop codon given"  
XX  
FN WO2004058969-A1.  
XX  
PD 15-JUL-2004.  
XX  
PE 22-DEC-2003; 2003WO-JP016417.  
XX  
PR 24-DEC-2002; 2002JP-00373144.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Hikichi Y, Nishizawa S;  
XX  
DR WI; 2004-525885/50.  
DR P-PSDB; ADP95920.  
XX  
PT Novel prophylactic and therapeutic agent of cancer, having compound or  
PT its salt that inhibits activity of enhancer of zeste Drosophila homolog2  
PT protein, useful for treating cancer such as lung cancer, breast cancer.  
XX  
PS Disclosure; SEQ ID NO 2; 102pp; Japanese.  
XX  
CC The invention relates to an agent for the prophylaxis and therapy of  
CC cancer comprising a compound which inhibits the activity or expression of  
CC human E2H2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
CC E2H2 inhibitors include anti-E2H2 antibodies and E2H2 antisense nucleic  
CC acids. The invention also relates to a kit for screening for inhibitors  
CC of E2H2 activity or expression. E2H2 is a component of a complex with  
CC histone methyltransferase activity which methylates histone H3 on lysine  
CC residues at position 9 and/or 27. It has been found to be deregulated in  
CC various tumours, and acts as a mediator of tumorigenesis, being able to  
CC reverse repression of cyclin A which in turn mediates cell cycle  
CC advancement. The agent of the invention can be used for preventing,  
CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
CC in drug screening. The agent of the invention is safe and less toxic than  
CC prior art cancer prophylactic/therapeutic agents. The present sequence

CC represents DNA encoding human EZH2.

XX Sequence 2253 BP; 717 A; 449 C; 546 G; 541 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	7,42e-281	Length:	2253
Score:	3787.50	Matches:	698
Percent Similarity:	99.1%	Conservative:	0
Best Local Similarity:	99.1%	Mismatches:	1
Query Match:	99.5%	Indels:	5
DB:	12	Gaps:	1

US-10-773-302-2\_COPY\_1\_699 (1-699) x ADP95921 (1-2253)

```

QY      1 MetGlyGlnThrGlyLeuLeuSerGlyLeuGlyProValCysThrArgLeuValLeuLeu
DB      1 AAGGGCCAGACTGGGAGAAATCTGAGAAAGGAGCCGTTGTGGGAGAAAGGCTGTAATA
        60
QY      21 SerGluTyrMetArgLeuArgGlnLeuLeuSerArgPheArgAlaAspGluValLeuSer
DB      61 TCGAGTACATCGACTGAGACAGCTCAGAGGCTTCAAGCTGATGATGAAGTAAAGCT
        120
QY      41 MetPheSerSerAsnArgGlnLeuLeuGlnArgThrGlnLeuAsnGlnGlnTyr
DB      121 ATGTTAAGTTCATTCGTCAAAAATTTTGAAAAGACGGAATCTTAAACCAAGATCG
        180
QY      61 LysGlnArgArgGlnGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr
DB      181 AAACGCGAAGATACAGCTGTGACATCTGACCTTGTGAGCTCATTTGCGCGGACT
        240
QY      81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLeuThr
DB      241 ACGAGGTTCGCTGACGACGATGATTTTCAACACAGTCAATCCATTAAAGACT
        300
QY      101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnIleAsn
DB      301 CTGAATGAGTTCCTTCAAGTACCATTAATGTATTTCTGTGCTCCCTTACAGCAATTTT
        360
QY      121 MetValGluAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAsp
DB      361 ATGGGGAAGATGAACCTGTTTACATTAATCTCTTATATGGAGATCAAGCTTTTATGAT
        420
QY      141 GlnAspGlyThrPheIleGlnGlnLeuLeuLeuAsnTyrAspGlyLeuValHisGlyAsp
DB      421 CAGGATGGTACTTCACTTGAAGAACCTAAATAAATATGATGGAGAAAGTACACGGAGAT
        480
QY      161 ArgGluCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlnGln
DB      481 AGAGATGTGGCTTATATAATGATGAATTTTGTGGAGTTGTGAATGCCCTTGTCTCA
        540
QY      181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGlnArgGlnGlnGln
DB      541 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
        600
QY      201 LysAspAspLeuGlnAspHisArgAspAspLeuSerArgProProArgLeuLeuPhePro
DB      601 AAAGATCTGAGAGATCAACGAGATGATTAAGAAACCGCCACCTCGGAATTTCTCTCT
        660
QY      221 AspLeuIlePheGlnAlaIleSerSerMetPheProAspLeuGlyThrAlaGlnGlnLeu
DB      661 GATTAATAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGCGACAGCAAGAAACTA
        720
QY      241 LysGlnGlyTyrLeuGlnLeuThrGlnGlnGlnLeuProGlyValIleLeuProGlnGln
DB      721 AAGGAAAAATATTAAGAACTCAACGACAGCTCCGAGCGGCGCACTTCTCTCAAGCT
        780
QY      261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer
DB      781 ACCCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
        840
QY      281 PheHisThrLeuPheCysArgArgCysPheLeuHisIle----- 297
DB      841 TTTGATAGGCTTTTCTGTAGGCGATGTTTAAATATGATGATGATGATGATGATGATGAT
        900

```

```

QY      298 -----ProPheHisAlaThrProAsnThrTyrLeuArgLeuAsnThrGlnThrAlaLeu
DB      901 AATTATTCCTTTTCAAGAACACCCAACTTATTAAGGAGACACAGAAACAGCTCTTA
        960
QY      316 AspAsnLeuProCysGlyProGlnCysTyrGlnHisLeuGlnGlnValAlaLeuGlnPheAla
DB      961 GACAAACAAACCTTGTGGACCAACAGTGTATACAGCATTTTGGAGGAGCAAAAGAGTTTGTCT
        1020
QY      336 AlaAlaLeuThrAlaGlnArgGlnLeuLeuThrProProLeuSerProGlnGlyArgArgArg
DB      1021 GCTGCTTCACCGCTGAGCGGATTAAGACCCCAACCAAAAGCTCCAGAGGCGCGAGAAAGA
        1080
QY      356 GlnArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGlnLeuSer
DB      1081 GAGCGGCTTCCCAATTAACATGATGACAGGCGCCAGACCCCACTTAATATGTGCTGAATCA
        1140
QY      376 LysAspThrAspSerAspArgGlnAlaGlyThrGlnArgGlnGlnValAsnAsnAspLeu
DB      1141 AAGGATACAGACAGTATAGGGAACAGGACTGAAGCGGGGAGAGAACATGATATAA
        1200
QY      396 GlnGlnGlnGlnLeuLeuAspGlnThrSerSerSerSerSerGlnAlaAsnSerArgCysGln
DB      1201 GAAGAAAGAAAGAAAGAAAGAACTTCAGACTCTCTGAAAGCAAAATTCGCTGTCAA
        1260
QY      416 ThrProIleLeuMetLeuProAsnIleGlnProProGlnAsnValGlnTyrSerGlyAla
DB      1261 ACACCAATTAAGATGAACCAAAATTAATTAACCTCTGAAATGTGAGATGAGTGTGTGT
        1320
QY      436 GlnAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla
DB      1321 GAAGCTCAATGTTTAGTGTCTTCACTTGGCACTTACTATGACAAATTTCTGTGCCATTGCT
        1380
QY      456 ArgLeuIleGlyThrLeuThrCysArgGlnValTyrGlnPheArgValLeuGlnSerSer
DB      1381 AGGTTAATTTGGGACCAAAACATGTAGACAGGTGTATGATTTAGTCAAGAAATCTAGC
        1440
QY      476 IleIleAlaProAlaProAlaGlnAspValAspThrProProArgGlyLeuLeuArgGly
DB      1441 ATCATAGCTCAAGCTCCCGTGAGGATGTGATATCTCTCAAGGAAAGAAAGAGGAA
        1500
QY      496 HisArgLeuTyrPheAlaHisCysArgGlyIleGlnLeuLeuLeuAspGlySerSerAsn
DB      1501 CACCGGTTGTGGCTGCACATCGCAGAAAGATACAGCTGAAAGAGAGCGCTCTCTTAAC
        1560
QY      516 HisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysPro
DB      1561 CATGTTTACACTATCAACCTCGATCATCAACGCGAGCGCTTGTGACAGTTCCGCGCCT
        1620
QY      536 CysValIleAlaGlnAsnPheCysGlnLeuPheCysGlnCysSerSerGlnCysGlnAsn
DB      1621 TGTGTGATGACAAATAATTTTGTGAAAGTTTGTCAATGTATGATTCAGAGTGTCAAAAC
        1680
QY      556 ArgPheProGlyCysArgCysLeuAlaGlnCysAsnThrLysGlnCysProCysTyrLeu
DB      1681 CGCTTTCCGGAATGCCCTGCAAGACAGTGCACCAACCAAGAGTCCCTGTACTCTG
        1740
QY      576 AlaValArgGlnCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIleThrAsp
DB      1741 GCTGTCGAGAGTGTGACCTGACCTGTCTTACTTGTGAGCGCGTGCACATTTGGGAC
        1800
QY      596 SerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeu
DB      1801 AGTAAAAATGTGCTCTGCAAAACCTGCAAGTTCAGCGGCGCTCCAAAAACATCTATTTG
        1860
QY      616 LeuAlaProSerAspValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsn
DB      1861 CTGGCACTATCTGACGTGGCGAGCTGGGGAATTTTATCAAAAGTCTGTGCAAAAAT
        1920
QY      636 GlnPheIleSerGlnTyrCysGlyGlnIleIleSerGlnAspGlnAlaAspArgArgGly
DB      1921 GAATTCATCTCAGATATCTGTGAGAGATTTATTTCTCAAGATGAAGCTGACAGAAAGGG
        1980

```

QY 656 LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheValVal 675  
DB 1981 AAAGTATGATTAATATACATGCGCTTTCTGTTCACTTGAACAAATGATTTTGTGGTG 2040  
QY 676 AsparAlaThrArgLysGlyValAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695  
DB 2041 GATGCAACCCCGCAAGGTTACAAATTCGTTTGGCAATTCATTCGGTAAATCCAAATCGC 2100  
QY 696 TyrAlaLysVal 699  
DB 2101 TATGCAAAAGTT 2112  
RESULT 15  
ADP95922  
ID ADP95922 standard; DNA; 2695 BP.  
XX  
AC ADP95922;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human EZH2 (enhancer of zeste homologue 2)-encoding DNA, SEQ ID NO:3.  
XX  
KW Human; EZH2; enhancer of zeste homologue 2; chromosome 7q35;  
KW histone methyltransferase; histone H3; cell cycle; tumorigenesis;  
KW inhibitor; apoptosis; cancer; colorectal cancer; breast cancer;  
KW lung cancer; pancreatic cancer; ovarian cancer; drug screening;  
KW prophylaxis; prevention; diagnosis; cytostatic; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH CDS 167..2422  
FT /\*tag= a  
FT /product= "EZH2 (enhancer of zeste homologue 2)"  
FT  
XX MO2004058969-A1.  
XX  
PD 15-JUL-2004.  
XX  
PP 22-DEC-2003; 2003MO-JP016417.  
XX  
PR 24-DEC-2002; 2002JP-00373144.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
XX Hikichi Y, Nishizawa S;  
XX PI WPI; 2004-525885/50.  
XX DR P-PSDB; ADP95920.  
XX  
PT Novel prophylactic and therapeutic agent of cancer, having compound or  
PT its salt that inhibits activity of enhancer of zeste Drosophila homologue2  
PT protein, useful for treating cancer such as lung cancer, breast cancer.  
XX  
XX Example 3; SEQ ID NO 3; 102pp; Japanese.  
XX  
CC The invention relates to an agent for the prophylaxis and therapy of  
CC cancer comprising a compound which inhibits the activity or expression of  
CC human EZH2 (enhancer of zeste homologue 2), resulting in apoptosis. The  
CC EZH2 inhibitors include anti-EZH2 antibodies and EZH2 antisense nucleic  
CC acids. The invention also relates to a kit for screening for inhibitors  
CC of EZH2 activity or expression. EZH2 is a component of a complex with  
CC histone methyltransferase activity which methylates histone H3 on lysine  
CC residues at position 9 and/or 27. It has been found to be deregulated in  
CC various tumours, and acts as a mediator of tumorigenesis, being able to  
CC reverse repression of cyclin A which in turn mediates cell cycle  
CC advancement. The agent of the invention can be used for preventing,  
CC treating and diagnosing cancers such as colorectal cancer, breast cancer,  
CC lung cancer, pancreatic cancer and ovarian cancer, and can also be used  
CC in drug screening. The agent of the invention is safe and less toxic than  
CC prior art cancer prophylactic/therapeutic agents. The present sequence  
CC represents a DNA encoding human EZH2 which was used in an example of the  
CC invention.

XX  
SQ Sequence 2695 BP; 839 A; 549 C; 660 G; 647 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 9,37e-281 Length: 2695  
Score: 3787.50 Matches: 698  
Percent Similarity: 99.1% Conservative: 0  
Best Local Similarity: 99.1% Mismatches: 1  
Query Match: 99.5% Indels: 5  
DB: 12 Gaps: 1  
US-10-773-302-2\_COPY\_1\_699 (1-699) x ADP95922 (1-2695)  
QY 1 MetGlyGlnThrGlyLysSerGlyLysGlyProValCysTrpArgLysValLys 20  
DB 167 ATGGCCGACACTGGAGAAATCTGAGAAAGGAGCACAGTTGTTGGCGGAAGCGGTAA 226  
QY 21 SerGluTyrMetArgLysValGlyGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 227 TCAGAGTACATGCACTGAGACAGCTCAAGGGTTCAAGACAGCTGATGAATGAAGGT 286  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnLysTrp 60  
DB 287 ATGTTTACGTTCCATCGTCAGAAATTTGGAAAGAACGAAATCTTAACCAAGAAATGG 346  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACAGCGAAGATACAGCTCTGACATCTCTGACCTCATTTGCGGAGACT 406  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 407 AGGAGGTTCGGAGCCAGTACGACTGGATTTCCACACAGTCACTCCATTAAAGACT 466  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB 467 CTGAATGAGTTGCTTCAATGACCAATATGATTTCTTGCTCCCTTACAGCAAAATTT 526  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 527 ATGGTGGAAAGATGAAACTGTTTACATPAACATTCCTTAATAGGAGATGAAGTTTGTAGT 586  
QY 141 GlnAspGlyThrPheIleGluGlnLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 587 CAGATGTACTTTCATTTGAAGAACTAATTAATAAATTAATGAGGAAGTCAAGGGAT 646  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 647 AGAGAAATGCGTTTATTAATGATGAATTTTGTGAGTTGGTGAATGCCCTTGTCTCA 706  
QY 181 TyrAsnAspAspAspAspAspAspAspAspGlyAspAspProGluGlnArgGlnLysGln 200  
DB 707 TATATATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
QY 201 LysAspLeuGlnGluAspHisArgAspAspLysGluSerArgProArgLysPheProSer 220  
DB 767 AAAGATTCGAGAGATCACCGAGATGATTAAGAAAGCGCCCACTCGGAAATTTCTTCT 826  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGlnLeu 240  
DB 827 GATTAATAATTTTGAAGCAATTTTCCCAATGTTTCCAGATTAAGGCGACAGCAAGAACTA 886  
QY 241 LysGluLysTyrLysGlnLeuThrGluGlnGlnLeuProGlyAlaLeuProGluCys 260  
DB 887 AAGGAAAAATTAAGAACTACCGAAGACAGAGCTCCCGAGCGGCACTTCTCCGAAGGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer 280  
DB 947 ACCCCCAACATAGAGTGAAGCAAAATGCTAAATCTGTTCAAGAGAGCAAAAGCTTACACTCC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHis----- 297  
DB 1007 TTTCAATACGCTTTTCTGTAGCGATGTTTAAATATGACTGCTTCTTACATCGTAAGTGC 1066

QY 238 -----ProPheH1eAlaThrProAsnThrTyLysArgLysAsnThrGluThrAlaLeu 315  
DB 1067 AATTATCTTTTCATGCAACACCCAACTTTATTAAGGAGAAACACAAACAGCTCTA 1126  
QY 316 AAPAanLysProCysGlyProGlnCysTyArgInH1sLeuGluGlyAlaLysGluPheAla 335  
DB 1127 GACAAACAACCTTGTGACACAGGTGTTCACAGCATTTGAGAGGACAAAGAGTTTGTCT 1186  
QY 336 AlaAlaLeuThrAlaGluArgLysLeuThrProProLysArgProGlyGlyArgArgArg 355  
DB 1187 GCTGCTTCACCGCTGAGCGGATTAAGACCCCAACAAAGCTCCAGAGCGCGCAGAAAC 1246  
QY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrL1eAsnValLeuGlnSer 375  
DB 1247 GAAAGGCTTCCCAATACAGTACAGAGCCCAAGCACTTATATATGCTGAGATCA 1306  
QY 376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyLysAsnAsnAspLys 395  
DB 1307 AAGGATACAGACAGATGAGGAGAGGAGGAGCTGAAGCGGGGAGAGAAACATGATATA 1366  
QY 396 GluGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415  
DB 1367 GAAGAGAGAGAGAGAAAGATGAACCTTGAGCTCTCTGAAGCAAAATTCGGTGTCAA 1426  
QY 416 ThrProL1eLysMetLysProAsnL1eGluProProGluAsnValGluThrSerGlyAla 435  
DB 1427 AACCAATTAAGATGAAGCCAAATTAATTAACCTCTGAGAAATGAGAGTGGTGTCT 1486  
QY 436 GluAlaSerMetPheArgValLeuL1eGlyThrTyTyAspAsnPheCysAlaL1eAla 455  
DB 1487 GAAGCTCAATGTTTAGAGTCTCTCATTTGCACTTACTATGACAAATTCGTGCAATGCT 1546  
QY 456 ArgLeuL1eGlyThrLysThrCysArgGlnValTyGluPheArgValLysGlnSerSer 475  
DB 1547 AGGTTAATTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAGACTCAAGAATCTAGC 1606  
QY 476 L1eL1eAlaProAlaProAlaGluAspValAspThrProArgLysLysLysArgLys 495  
DB 1607 ATCATAGCTCCAGCTCCGCTGAGATGTGATACCTCTCAAGGAAAGAGAGAGAA 1666  
QY 496 HisArgLeuThrAlaAlaHisCysArgLysL1eGlnLeuLysLysAspGlySerSerAsn 515  
DB 1667 CACCGGTTGTGGCTGCACACCTGCAGAAAGATACAGCTGAAGGAGGAGCGCTCTTAAC 1726  
QY 516 HisValTyArgInTyArgInProCysAspHisProArgGlnProCysAspSerSerCysPro 535  
DB 1727 CATGTTTCAACTATCAACCTGTGATCATCCAGCGGAGCTTGTGACAGTTCTGTGCCCT 1786  
QY 536 CysValL1eAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsn 555  
DB 1787 TGTGTGATAGCAACAATTTTGTGAAAAGTTTGTCAATGTAGTTCAAGATGTCAAAAC 1846  
QY 556 ArgPheProGlyCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyLeu 575  
DB 1847 CGCTTCCGGGATGCGCTGCAGAAACAGTCAACACCAAGCAGTCCCGCTGCTAACC 1906  
QY 576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAsp 595  
DB 1907 GCTGTCCGAGAGTGAACCTGACCTGTCTTACTGTGAGCGGCTGACCATTTGGGAC 1966  
QY 596 SerLysAsnValSerCysLysAsnCysSerL1eGlnArgGlySerLysLysHisLeuLeu 615  
DB 1967 AGTAATAAAGTGTCTGCAGAAACAGTCAATTCAGGGGGCTCCAAAAAGCATCTATTG 2026  
QY 616 LeuAlaProSerAspValAlaGlyTrpGlyL1ePheL1eLysAspProValGlnLysAsn 635  
DB 2027 CTGGCAACATCTGAGGTGCAAGCTGGGGGATTTTATCAAAAGATCTGTGCAAAAAT 2086  
QY 636 GluPheL1eSerGluTyTyCysGlyGluL1eL1eSerGlnAspGluAlaAspArgArgGly 655  
DB 2087 GAATTCATCTCAGAAATACGTGAGAGATTATTTCTCAAGATGAAGCTGACAGAAAGGG 2146  
QY 656 LysValTyAspLysTyMetCysSerPheLeuPheAsnLeuAsnAspPheValVal 675

DB 2147 AAAGTATGATTAATATACATGTCAGCTTCTGTTCAACTGACATGATTTGTGTG 2206  
QY 676 AAPAlaThrArgLysGlyAsnLysL1eArgPheAlaAsnHisSerValAsnProAsnCys 695  
DB 2207 GATGCAACCCGCAAGGTTAACAAATTCGTTTGCAAATCATTCGGTAAATCCAACATGC 2266  
QY 696 TyAlaLysVal 699  
DB 2267 TATGCAAAAGTT 2278

Search completed: August 13, 2006, 13:50:03  
Job time : 1126.99 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 13, 2006, 13:12:07 ; Search time 6912.6 Seconds

(without alignments)  
8481.816 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808  
Sequence: 1 MGQTKSEKBPVCMKRRVK.....KGNKIRFANHSVNPYCAKV 699

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/abs/ABSSWMB spool/US10773302/runat\_11082006\_140227\_16655/app\_query.fasta\_1  
-DB=EST -OPMT=fastad -SUPFIX=p2n.rst -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs05p  
-USBR=US10773302 @CG1\_1\_12176 @runat\_11082006\_140227\_16655 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_est7: \*  
7: gb\_est8: \*  
8: gb\_est9: \*  
9: gb\_est10: \*  
10: gb\_est11: \*  
11: gb\_est12: \*  
12: gb\_est13: \*  
13: gb\_est14: \*  
14: gb\_est15: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3747	98.4	2652	AK157853 Mus muscu
2	3204	84.1	2241	AK157853 Mus muscu
3	3157	82.9	2241	AK157853 Mus muscu
4	2589	66.0	2241	AK157853 Mus muscu

5	2389.5	62.7	4089	AK154565 Mus muscu
6	2389.5	62.7	4194	AK140694 Mus muscu
7	2381.5	62.5	4643	CR857825 Pongo pyg
8	2273.5	59.7	2142	AK138942 Mus muscu
9	2245.5	59.0	2215	AY402398 Mus muscu
10	2243.5	58.9	2244	AY402396 Homo sapi
11	2157	56.6	2183	AY402397 Pan trogl
12	2048	53.8	1238	AK153835 Mus muscu
13	1645.5	43.2	2860	AK164192 Mus muscu
14	1645.5	43.2	3639	AK045374 Mus muscu
15	1579	41.5	873	CR990276 CR990276
16	1568	41.2	967	BQ946546 AGENCOURT
17	1557	40.9	859	BM974704 BM974704
18	1538	40.4	837	DR762983 DR762983
19	1538	40.4	844	CX514114 JGI XZG44
20	1527	40.1	972	BUS56770 AGENCOURT
21	1517	39.8	840	CX783825 HESG3_26
22	1507	39.6	842	CX870387 HESG4_44
23	1504	39.5	868	CX466741 JGI XZG47
24	1503.5	39.5	971	CO647894 ILLUMIN
25	1499	39.4	956	BQ653044 AGENCOURT
26	1497	39.3	1019	BQ065838 AGENCOURT
27	1491	39.2	828	DN096340 JGI CABR6
28	1474	38.7	869	DV927839 LB03011.C
29	1459	38.3	797	CX957081 JGI CAA09
30	1449	38.1	1204	DN702484 CLJ41-H09
31	1446	38.0	828	DV868931 LB0251.CR
32	1442	37.9	804	CX387198 JGI XZT21
33	1437.5	37.7	965	BQ654181 AGENCOURT
34	1426	37.4	792	BUS32465 B03498079
35	1425	37.4	946	BR729850 601564771
36	1417.5	37.2	890	BQ24971 AGENCOURT
37	1416.5	37.2	798	BUS30050 603117150
38	1416.5	37.2	904	CA454418 AGENCOURT
39	1411	37.1	876	BM040808 60314275
40	1401	36.8	840	DR845527 JGI CABR1
41	1397	36.7	967	BQ649256 AGENCOURT
42	1396	36.7	884	CY855194 gonad EST
43	1383	36.3	870	BX462271 BX462271
44	1382	36.3	982	BUS15131 AGENCOURT
45	1377	36.2	842	CF148364 AGENCOURT

#### ALIGNMENTS

RESULT 1  
AK157853  
LOCUS  
DEFINITION  
MUS musculus erythroblast cDNA, RIKEN full-length cDNA library,  
full insert sequence.  
ACCESSION  
AK157853  
VERSION  
AK157853.1 GI:74186125  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2  
3  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3  
Shibata, K., Itoh, M., Mizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,



Komuro, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Itoh, K., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, S., Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Giers, C., Kling, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G., Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagnier, L., Washio, T., Sakai, K., Okido, T., Furumoto, A., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotam, M., Hume, D.A., Kamya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazza, R., Mombauer, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyshew-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y.

CONSRMT  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
11217851  
Okazaki, Y., Furumoto, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Chotla, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chotla, C., Corbani, L.E., Cousins, S., Dalla, B., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kana, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Meglath, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numa, K., Okido, T., Pavan, M.J., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B., Ringwald, M., Sandelin, A., Schneider, C., Semp, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshew-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayashizaki, Y., Hirokawa, N., Kawai, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., and Hayashizaki, Y.

CONSRMT  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)  
12466851  
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Fritth, M.C., Maeda, N., Oyama, R., Ravasi, T., Lennard, B., Wells, C., Kodling, R., Shimokawa, K., Bajic, V.B., Bremner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, V., Allen, J.E.,

CONSRMT  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)  
16141072  
Katsuyama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yagi, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Fritth, M., Ravasi, T., Pang, K.C., Hallinan, J., Matlick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Peghini, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lennard, B., and Wahlestedt, C.

CONSRMT  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Riken Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)  
16141073  
8 (bases 1 to 2652)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imoto, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Komuro, H., Murata, M., Nakamura, M., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Science Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukuba, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Science Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Hiromitsu Nakauchi (Center for Experimental Medicine, Institute of Medical Science, University of Tokyo 4-6-1, Shirokane-dai, Minato-ku, Tokyo, 108-8639, JAPAN) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>

## FEATURES

Location/Qualifiers

source

1..2652

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM\_DB:K0C0026809"

/db\_xref="taxon:10090"

/clone="K0C0026809"

/cell\_type="erythroblast"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

CDS

146..2386

/note="unnamed protein product, enhancer of zeste homolog 2 (Drosophila) (MGD|WGI:107940 GB|NM\_007971, evidence: BLASTN, 99%, match=2635)"

/putative"

/codon\_start=1

/protein\_id="BAB34232.1"

/db\_xref="GI:74186126"

/translation="MGQTKSKSEKGPVCMKKVKSEYMKLRQLKPRADAEVKTMS

NROKILERTETLQEMKORIQHVIHMTSVSLRGTRCSTVDLPAQVPLKTLN

AVASVIMYSNPLQONPMVEDETVIMTIPYWGDEVLDQDGFIELIKNDGKVG

RRCGFINDIPELVNAGVYNDDDDDDDPDRHKKKOLBNRDKETCPKRP

PADKIPKAISSMPPKGTABEKYKELTEQLPGALPRECTPIIDPNAKVORQ

SLHSFTLPCRRCFRTDCLPHFPAFTVTKTKNTETLNDKPCGQCTOHLEKAE

AAALTAERIKTTPKRGRRGRRLPNNSRPSTPTISLVESKDTSDREAGTETGEN

## Alignment Scores:

Pred. No.: 0 Length: 2652  
Score: 3747.00 Matches: 686  
Percent Similarity: 99.1% Conservative: 7  
Best Local Similarity: 98.1% Mismatches: 6  
Query Match: 98.4% Indels: 0  
DB: 6 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x AK157853 (1-2652)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB 146 ATGGCCAGACTGGGAGAAATCTGAGAGGAGCCGGTTGTTGGCGGAAAGGTGTAA 205  
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
DB 206 TCAGAGTACATGAGCTGAGCAGCTCAAGAGTTCAGAGAGCTGATGATTAAGACT 265  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGlnIleLeuAsnGlnGlnUtrp 60  
DB 266 ATGTTAGTCCAAATCGTCAAGAAATTTGGAAAGAACTGAAACCTTAACCAAGAGTGG 325  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 326 AAGCAGCGAGATACAGCATGTCACATCCTTCGTGAGCTCANTTGGCGGACT 385  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 386 AGGGAGTTCAGTCAACGATGATTTTCCAGCAGAGTCAATCCGTTAAAGACC 445  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB 446 CTGAATGAGTGCCTCGGTGCTTAATAGTACTCTGTGCGCTTACACAGAAATTT 505  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 506 ATGGTGGAGAGGAAAGCTTTTACATTAATCTCTTAATATGGGGAGTGAAGTTCGAT 565  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 566 CAGGATGGCACTTTCATGAGAACTAATTAATTAATGATGAGAAAGTCAATGATGAC 625  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnGln 180

DB 626 AGAGATGTGATTTATTAATGATGAAATTTTGTGAGTGTGTAAATGCTTGTGTCA 685  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
DB 686 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 745  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerAspProProArgLysPheProSer 220  
DB 746 AAGATCTAGAGATTAATCGAGATTAAGAAACTTCCCACTCGAATAATTTCTGT 805  
QY 221 AspLysIlePheGluAlaIleSerSerPheProAspLysGlyThrAlaGluGluLeu 240  
DB 806 GATTAATATTTTAAGGCAATTTCTCAATGTTTCCAGATTAAGGCGACCGAGAAAGACTG 865  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGluLeuProGlyAlaLeuProProGluCys 260  
DB 866 AAGGAAATATTAAGAACTCAAGAGAGAGAGCTCCAGAGTCTCTCTCTTAATGT 925  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 926 ACTCAACATCGATGAGACCAATGCAATCTGTTCAGAGGAGCAAGCTTGATTC 985  
QY 281 PheIleThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisPhePheHis 300  
DB 986 TTTGATACGCTCTTCTGTGCAAGATGTTTAAGATGACTGCTTCTCATTCCTTCAT 1045  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 1046 GCACACCCACACATATTAAGAGAGAAACACAGAAAGCTTTGAGACAAAGCTTGT 1105  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGlyAlaLysGluPheAlaAlaLeuThrAla 340  
DB 1106 GAGCAGAGTGTATACACACTGAGAGGAGCTTAAGAGTGTGTCTGTCTTACTGT 1165  
QY 341 GluArgIleLysThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360  
DB 1166 GAGGCTATTAAGACACCACTTAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1225  
QY 361 AsnSerSerAspProSerThrProThrIleAsnValLeuGluLysLysAspThrAspSer 380  
DB 1226 AACGTACAGACCCAGACCCCACTCATGATGTCGAGTCAAGATACAGACAGT 1285  
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAsnAspLysGluGluGluLys 400  
DB 1286 GACAGAGAGGAGGAGCTGAATCGGGAGAGAGCAATGATTAAGAGAGAGAGAGAA 1345  
QY 401 LysAspGluThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1346 AAGATGAGAGCTCAGCTCTCTGACCAAAATCTCGGTGCAACCAATTAAGATG 1405  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1406 AAGCAAAATATTAAGCTCTGAGAGATGAGTGAATGAGTGAAGCTTCAATGTTT 1465  
QY 441 ArgValLeuIleGlyThrTyrThrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1466 AAGTCTCATATGACTTACTACATTAATCTTGTGTGCAATGCTGAGCTTAATGAGACC 1525  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1526 AAAACATTAAGACAGTGTATGAGTTTAGAGTCAAGAGTCCAGATCATAGCACTGT 1585  
QY 481 ProAlaGluAspValAspThrProProArgLysLysValGlyValHisArgLeuThrAla 500  
DB 1586 CCCACTGAGATGTAAGACTCTCTCAAGAAAGAGAAAGAAACATCGGTGAGGCT 1645  
QY 501 AlaHisCysArgLysIleGluLeuLysValAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1646 GACACTGACAAAGATACATGAAAGAGAGAGGCTCTCTTACCAATGTTTACACTAT 1705  
QY 521 GlnProCysAspHisPheProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1706 CAACCTGTGACATCCAGGAGGCTTGTGACAGTGTGTGCTGTGTATACACAA 1765

QY 541 AsnPhcYcgluyluyspneCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1766 AATTGTTGTGAAAAGTTTGTGCATGTAGTTCAAGGTGTCAAAACCGCTTTCGTGAGT 1825  
QY 561 ArgCyslysalaglnCysAsnThrlyGlnCysProCyslyTleuAlaValArgGluCys 580  
DB 1826 CGGTGCAAGACATGACATGACACCAACAGGTGTCTATCTACCTGCTGCGAGAGGT 1885  
QY 581 AspProAspLeuCysLeuThrCysGlyValAlaAspHisTyrPaspSerLysAsnValSer 600  
DB 1886 GACCTGACCTCTGTCTCACTGTGAGACTGTCTGACCACTGGACAGTAAAAATGTATCC 1945  
QY 601 CyslyAsnCysSerIleGlnArgGlySerlylybHisleuLeuAlaProSerAsp 620  
DB 1946 TGTAAAGACTGTAGATTCAAGGGGGCTTAAAAAGCACTTACTGCTGGACCGTCTGAT 2005  
QY 621 ValAlaGlyTTPGlyIlePheIleLysAspProValGlnlybAsnGluPheIleSerGlu 640  
DB 2006 GTGGCAGGCTGGGGCATCTTATCAAAAGATCTGTACAGAAAAATGAATTCATCTCAGAA 2065  
QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgGlyLysValTyrAspLys 660  
DB 2066 TACTGTGGAGATTATTTCTCAGATAGACAGACAGAGAGAAAAAGTGTATGACAAA 2125  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
DB 2126 TACATGTGCACTTTCTGTCACTTGAACATGATTTTGTGTGAGATGCAACCGGAAAG 2185  
QY 681 GlnAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
DB 2186 GGCAACAAATTCGTTTGTCTAATCATTCAGTAATCCAACTGCTATGCAAAAGTT 2242

RESULT 2  
LOCUS AY399441 2241 bp DNA linear GSS 15-DEC-2003  
DEFINITION Homo sapiens EHZ2 gene, VIRUTAL TRANSCRIPT, partial sequence,  
ACCESSION AY399441  
VERSION AY399441.1 GI:39755430  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukariyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 2241)  
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
Science 302 (5652), 1960-1963 (2003)  
14671302  
2 (bases 1 to 2241)  
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
SOURCE  
Location/Qualifiers  
1..2241  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>2241  
/gene="EHZ2"  
/locus\_tag="HCM0225"

ORIGIN  
Alignment Scores:  
Pred. No.: 1,12e-286 Length: 2241  
Score: 3204.00 Matches: 601  
Percent Similarity: 89.6% Conservative: 0  
Best Local Similarity: 89.6% Mismatches: 70  
Query Match: 84.1% Indels: 0  
DB: 14 Gaps: 0  
US-10-773-302-2\_copy\_1\_699 (1-699) x AY399441 (1-2241)  
QY 1 MetGlyGlnThrGlyLyslybSerGluLysGlyProValCysTyrArgLysArgValLys 20  
DB 1 ATGGGCCAGACTGGGAAAGAAATCGAAGAGGACCAATTTGTTGGCGGAAACGTGAAA 60  
QY 21 SerGluTyrMetArgLeuArgGlnLysLysArgPheArgAlaAspGluValLysSer 40  
DB 61 TCAGAGTACATGCACTGACAGACACTCAAGAGGTTTCAGACGACTGATGAAGTAAAGACT 120  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTyr 60  
DB 121 ATGTTTAGTTCCATGCTGCAAGAAATTTGGAAAGACGGAAATCTTAAACCAAGATGG 180  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 181 AAACAGCGAAGATACAGCGCTGTGCACATCTGTCTGTGACCTCATTTGCGGGGACT 240  
QY 81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGluValIleProLeuLysThr 100  
DB 241 AGGAGAGNN 300  
QY 101 LeuAsnAlaValAlaSerValProIleMetLysSerTyrSerProLeuGlnAsnPhe 120  
DB 301 NNN 360  
QY 121 MetValGlnAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 361 NNNNGTGAGATGAAACGTGTTTACATACATCTCTATATGGAGATGAAGTTTATGAT 420  
QY 141 GlnAspGlyThrPheIleGlnGluLysIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
DB 421 CAGATGTACTTTTCATTGAAAGACTAATTAATAAATTTATGATGGAAAGTACAGGGGAT 480  
QY 161 ArgGlyCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGGTAATGCCCTTGCTGCA 540  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluLysGluGln 200  
DB 541 TATATATGATGATGACATGATGATGAGACATCTCGAAGAAAGAAAGAAAGCAG 600  
QY 201 LysAspLeuGluAspHisArgAspAspLysSerArgProArgLysPheProSer 220  
DB 601 AAAGATCTGAGAGATCACCGAGATGATGAAGAACCGGCCCACTCGGAATTTCTCTCT 660  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 661 GATAAAATTTTGAAGCATTTCTCAATGTTTCCATTAAGGGCAGCAGAGAAACTA 720  
QY 241 LysGluLysTyrLysGluLeuThrGlnGlnGlnLeuProGluValAlaLeuProProGluCys 260  
DB 721 AAGGAAAAATATTAAGAACTCACCGAAGACGAGCTCCAGCGGCACTTCTCGAAGT 780  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 781 ACCCCCAATGATGAGCAAAATGCTAATCTTTTCAAGAGAGCAAGCTTACATCC 840  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 841 TTTCATAGGCTTTCTGTAGGCGATGTTTAAATATGATGCTGCTTCTTAACNNNNNNNN 900  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320

[illegible][illegible]

[illegible]

QY	461	LysThrCysArgGlnValTyrGlnPheArgValIleGlnSerSerIleIleAlaProAla	480
Db	1381	AAAAcAGTGTACGACGGGTGTAGTGAATTGATGAGTCAAGAGTCCAGTATCATAGACCCCTGT	1440
QY	481	ProIaGlnuAPValIaAPThrProProArgIuYelYelYsYsArgIYsHsATLeuTPAla	500
Db	1441	CCCACTAGGATGTAGACACTCTCTCCAGAGAAAGAGAAAGAAACATTCGGTTGTGGCT	1500
QY	501	AlaHisCysArgIYsIleGlnLeuYsYsAPGlySerSerAsnHisValTyrAsnTyr	520
Db	1501	GCACACGCGAGAAAGATATCAACTGAAAAAGACGGCTCTCTAACCATGTTTCAACTAT	1560
QY	521	GlnProCysAspHisAPProArgGlnProCysAspSerSerCysProCysValIleAlaGln	540
Db	1561	CAACCTGTGTACATCCACCGCAGCCTTGAGACAGTTCCGTCCTGTGTGATAGACAA	1620
QY	541	AsnPheCysGluYsPhePheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys	560
Db	1621	AATTGTGTGAAAGTTTGTTCATGTATGTTCAAGTGTCAAAACCGCTTTCCTGATGT	1680
QY	561	ArgCysYsAlaGlnCysAsnThrIYsGlnCysProCysTyrLeuAlaValArgGluCys	580
Db	1681	CGGTGCAAAAGCAATGCAACACCAACAGTGTCACTACCGCTGTCCGAGAGTGT	1740
QY	581	AspProAspLeuCysLeuThrCysGlyValAlaAspHisSTTPAspSerIYsAsnValSer	600
Db	1741	GACCTTGACCTCTCTCTCAACGTGTGAGCGTGTACCATTTGGCACATTAATAAATGTATCC	1800
QY	601	CysIYsAsnCysSerIleGlnArgGlySerIYsYsHisIleuLeuAlaProSerAsp	620
Db	1801	TGTAAAGACTTATGACTTACCGGGGCTCTTAAAGACATTACTGTGGCACCGCTCGAT	1860
QY	621	ValAlaGlyTTPGlyIlePheIleYsAspProValGlnYsAsnGluPheIleSerGlu	640
Db	1861	GTGGAGGCTGTGGGCACTCTTATCAAGATCCGTACAGAAATGATTCATCTCAGAA	1920
QY	641	TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyYsValTyrAspIYs	660
Db	1921	TACTGTGGGAGATTATTTCTCAGAGTGAACGACAGACAGAAAGAGTATGACAA	1980
QY	661	TyrMetCysSerPheLeuPheAsnLeuAsn	671
Db	1981	TACATGTGACACTTCTGTTCACTTGAACAT	2013
RESULT 4			
AY399442		2241 bp	DNA linear
LOCUS	AY399442		SSS 15-DEC-2003
DEFINITION	Pan troglodytes EHz2 gene, VIRIDAL TRANSCRIPT, partial sequence,		
ACCESSION	AY399442		
VERSION	AY399442.1	GI:39755431	
KEYWORDS	SSS.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
REFERENCE	Bukacina, J., Metacosa, Chordata, Craniata, Vertebrata, Euteleostomi,		
AUTHORS	Mammalia, Eutheria, Euarchontoglires, Primates, Catarrhini,		
	Homnidae, Pan.		
	1 (bases 1 to 2241)		
	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,		
	Todd, M.A., Tenenbaum, D.M., Civeleo, D.R., Lu, F., Murphy, B.,		
	Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Smitsky, J.J.,		
	Adams, M.D. and Cargill, M.		
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous		
	gene trios		
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBMED	14671302		
REFERENCES	2 (bases 1 to 2241)		
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,		
	Todd, M.A., Tenenbaum, D.M., Civeleo, D.R., Lu, F., Murphy, B.,		
	Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Smitsky, J.J.,		
	Adams, M.D., and Cargill, M.		
TITLE	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,		
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,		

Rockville, MD 20850, USA  
This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES  
source  
1..2241  
Location/Qualifiers

gene  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..>2241  
/gene="BZH2"  
/locus\_tag="HCM0225"

## ORIGIN

Alignment Scores:  
Pred. No.: 2e-229 Length: 2241  
Score: 2589.00 Matches: 503  
Percent Similarity: 75.0% Conservative: 0  
Best Local Similarity: 75.0% Mismatches: 168  
Query Match: 68.0% Indels: 0  
DB: 14 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x AY399442 (1-2241)

QY 1 MetGlyGlnThrGlyLysLeuSerGlyLysGlyProValCysThrArgLysArgValLys 20  
DB 1 ATGGGCACAGACTGGAGAAATCTGAGAGGAGCAAGTTGTGGCGGAAAGCGTAAATA 60  
QY 21 SerGlyTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGlnValLysSer 40  
DB 61 TCAGAGTACATCGACTGAGACAGCTCAAGAGCTTCAAGAGCTGATGATTAAGANN 120  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGlnIleLeuAsnGlnGluTrp 60  
DB 121 NNN 180  
QY 61 LysGlnArgArgGlnGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 181 NNN 240  
QY 81 ArgGlyCysSerValThrSerAspLeuPhePheProThrGlnValIleProLeuLysThr 100  
DB 241 NNN 300  
QY 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB 301 NNN 360  
QY 121 MetValGlnAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAsp 140  
DB 361 NNNNGTGAAGATGAAGAACTGTTTACATTAATCTTATATGGAGATGAAGCTTTAGAC 420  
QY 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnTyrAspGlyLysValHisIleGlyAsp 160  
DB 421 CAGAGTGTACTTTCATTTGAGAACTAATTAATAAATTTATGATGAGAAAGTACACGGGAT 480  
QY 161 ArgGlyCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGln 180  
DB 481 AGAGNTGTGGGTTTATAAANGTGAATTTTGTGNNNTGTGAATNNNNNNNNNNNN 540  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnGluSerGln 200  
DB 541 NNN 600  
QY 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer 220  
DB 601 NNN 660  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGluLeu 240  
DB 661 GATATAAATTTTGAAGCACTTCTCAATGTTTCCAGATTAAGGCAACAGCAAGAAACTA 720  
QY 241 LysGlnLysTyrLysGlnLeuThrGlnGlnGlnLeuProGlyAlaLeuProProGlnCys 260

DB 721 AAGAAAAAANN 780  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnSerLeuHisSer 280  
DB 781 ACCCCCAACATNNNTGACCAAGCGTAACTCTGTCNNNNNNNNNNNNNNNNNNNNNNNN 840  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 841 TTTCATAGCTTTTNN 900  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnTrpAlaLeuAspAsnLysProCys 320  
DB 901 NNN 960  
QY 321 GlyProGlnCysTyrGlnHisLeuGlnGluValAlaLysGlnPheAlaAlaLeuThrAla 340  
DB 961 NNN 1020  
QY 341 GlnArgGlnIleLysThrProProLysArgProGlnGlyArgArgArgGlyArgLeuProAsn 360  
DB 1021 GAGCGGATTAAGACCCCAACAAAGCTCAAGAGGCGCGAGAGAGAGAGGCGCTTCCAAAT 1080  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1081 AACAGTACAGAGCCCAAGACCCCAACATTAATGCTGGAATCAAGATTAAGACAGT 1140  
QY 381 AspArgGlnAlaGlyThrGlnThrGlyGlnLysAsnAspLysGlnGlnGlnLys 400  
DB 1141 GATAGGAGAGAGGAGCTGAAGACGGGGGAGAGAAACAATGATTAAGAAAGAAAGAAAG 1200  
QY 401 LysAspGlnThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1201 AAAGATGAACCTTCAGCTCTCTGAGCAAAATTCCTGTCGTCGTCGTCGTCGTCGTCG 1260  
QY 421 LysProAsnIleGlnProProGlnLysAsnValGlnTrpSerGlnAlaGlnAlaSerMetPhe 440  
DB 1261 AAGCAATATATTGAACCTCTGAGAAATGTGAGATGAGATGAGATGAGATGAGATGAG 1320  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaLysLeuIleGlyThr 460  
DB 1321 AGAGCTCATATGAGCACTTACTATGACAAATTCCTGTCGTCGTCGTCGTCGTCGTCG 1380  
QY 461 LysThrCysAspGlnValTyrGlnPheArgValLysGlnSerSerIleAlaAlaProAla 480  
DB 1381 AAAACATGTAGACAGGTGTATGAGTTAGAGTCAAGAAATCTAGCATATAGCTCAGCT 1440  
QY 481 ProAlaGlnAspValAspThrProProArgLysLysValArgLysHisArgLeuTrpAla 500  
DB 1441 CCGCTGAGATGTGAGATCTCTCCAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1500  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1501 GCACATGCAAGAAAGATACAGCTGAAAAGACCGGCTCTCAACAGATTTACATATAT 1560  
QY 521 GlnProCysAspAspAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1561 CAACCTGTGTATCTCCAGCGAGCTTGTATCAGTTGTGTCCTGTGTGTATACACAA 1620  
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlnCys 560  
DB 1621 AATTTTGTGAAGTTTGTGTCATGTGTATGCTTCAAGATTCAAAACCGCTTCCGGGATTC 1680  
QY 561 ArgCysLysValGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGlnCys 580  
DB 1681 CGCTGCAAGACAGTGAACACCAAGAGAGCGGCTCTACCTGCTGCTGCTGCTGCTGCTG 1740  
QY 581 AspProAspLeuCysLeuThrCysGlnAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1741 GACCTTACCTCTGTCTTATCTGTGAGACCGCTCACTTGTGAGACAGTAAATATGTGCC 1800  
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAsp 620  
DB 1801 TCAGAAATCTGAGTATTCAGCGGGCTCCAAAAGCATCTATTCTGTCGACATCTGAC 1860

Oy	621	Val1AaglyTTPGlyIlePheIIeLysaApproValglnYsaangInupheIIesErglu	640	
Db	1861	GTGCACAGCGCTGGGGATTTTATTATCAAGAATCCTTGTCAGAAATAATGAATCATCTCAGAA	1920	
Oy	641	TyTcAgglYgIUlleIIesErglmaApGlnlaApPrargAgglyYvsVal1TyZAsplys	660	
Db	1921	TACGTGTGAGAGATTAATTTCTCAAGATCAACTGACAGAAAGGAGTATGATTA	1980	
Oy	661	TyMetCySserPheLeuPheAsnLeuasn	671	
Db	1981	TACATGTGCAGCTTTCTGTTCAACTGTACAAT	2013	
RESULT 5				
AK154565				
LOCUS				
DEFINITION	Mus musculus NOD-derived CD1ic +ve dendritic cells cDNA, RIKEN full-length enriched library, clone:F63004M06 product:enhancer of zeete homolog 1 (Drosophila), full insert sequence.	4089 bp mRNA linear HTC 21-SEP-2005		
ACCESSION	AK154565			
VERSION	AK154565.1	GI:74181971		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 10349636		
AUTHORS	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159		
JOURNAL	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, N., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoshida, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kir, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861		
REFERENCE	4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukumishi, Y., Kono, H., Akachi, O., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kanukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochwa, H., Kuehl, P., Lewis, S., Mateno, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabdi, F., Suzuki, R., Tomita, M., Wagner, D., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Cammici, P., de Bonaldo, M. F., Brownstein, M. J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Giustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P. P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombeerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitteker, C., Wilming, L., Wyman-Borje, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohzuki, S. and Hayashizaki, Y. RIKEN genome exploration Phase II Team and the PANOM Consortium		
CONSRITM				

FUNCTIONAL ANNOTATION OF THE MOUSE TRANSCRIPTOME BASED ON FUNCTIONAL ANNOTATION OF 60,770 FULL-LENGTH CDNAS	CONSRMT TITLE	JOURNAL PUBMED REFERENCE	AUTHORS
Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)			<p>5  Ozekaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Mikiado I., Otsu N., Saito R., Suzuki H., Yamana K., Miyazawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gotohori T., Baldirelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schmitt L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brade D., Brasic V., Chochua C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Fraser K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guetlich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenthart B., Lyons P.A., Maglott D.R., Maltas L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numa K., Okido T., Pavan W.J., Petosa G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Veardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilmberg L.G., Wyman-Boris A., Yangisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carrinci P., Hayatsu N., Hirozawa K., Shitaki T., Kono H., Nakamura M., Sakazawa N., Sato K., Shitaki T., Maki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.</p>
Functional annotation of the mouse transcriptome based on functional annotation of 60,770 full-length CDNAS			<p>6  Carrinci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic M.V., Brenner S.B., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilmberg L.G., Aldrich J., Allen J.E., Ambesi-Impolato A., Apweiler R., Aturaliya R.N., Bailey T.L., Banasal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Cline M.L., Dalla E., Dalrymple B.P., de Bono P., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Fukui S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gotohori T., Green R.E., Guetlich S., Harbers M., Hayashizaki Y., Hensch T.K., Hirokawa N., Hill D., Hummelbeck L., Iacono M., Ieko K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keisel J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J., Ljun S., McMillan S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone P., Miyake S., Moritz K., Mottaz-Fader S., Mulder N., Nakano N., Nakachi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Ozekaki Y., Orlandi V., Pang K.C., Pavan W.J., Pavoni G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Roest B., Ruan Y., Salzberg S.L., Sanderlin A., Schneider C., Schombach C., Sekiguchi K., Sempke C.A., Seno S., Smit C., Schombach C., Sekiguchi K., Sempke C.A., Seno S., Seese A., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sincalari B., Sperling S., Stupka E., Sugiyama K., Sultana R., Takanaka Y., Taki K., Tamajo K., Tan S.L., Tang S., Taylor M.S., Tenen J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamashita H., Zdobych E., Zhu S., Zimmer A., Hilde W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori Katayama M., Suzuki M., Aoki J., Aizawa T., Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Kono H., Nakano N., Nishio T., Okada M., Pleasy C., Shibata K., Shitaki T., Suzuki S., Tagami M., Waki K.,</p>





Db 1091 GAACCGTGTGGCAGACAGCTGCTCTTTGCTGGAAGAGCAAGATACGCC----- 1144  
 Qy 338 LeuThrAlaGluAlaGlyLeuThrProProlySerProGlyValArgArgGlyValArg 357  
 Db 1145 -----ATGCTGCACACCCCTCGCTCCAGAGCTCTGTGGCGCGCCGCCAAGG 1192  
 Qy 358 LeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu---GluSerLys 376  
 Db 1193 CACCAGAGGTGATGCTGCTCTCCAAATGCATCAGCTTCGCTATGCGTGAATACTAA 1252  
 Qy 377 AspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGlu 396  
 Db 1253 GAAGAGAGCAGGTGATGAGACACTGGC----- 1279  
 Qy 397 GluIuIuGlyLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGluThr 416  
 Db 1280 -----AATGACTGGGCGCTCCAGTCTTTCAGAGCTTAACTCTCGCTGTCAAGCC 1327  
 Qy 417 ProIleLysMetLysProAsn-----IleGlu---ProPro 427  
 Db 1328 CCCACGAAACGAAAGCCAGTCCAGCCCACTCAGCTCTGTGTGGTGAAGCCCTCA 1387  
 Qy 428 GluAsnValGluTrpSerGlyValGluAlaSerMetPheArgValLeuIleGlyThrTyr 447  
 Db 1388 GAGCCGGGTGAGTGGACCGGAGCCGAGAAATCTCTTCCAGTCTTCCACGAGCACTAT 1447  
 Qy 448 TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThrLysThrCysArgGluValTyr 467  
 Db 1448 TTCAACAACCTCTGCTCAATAGCCAGGCTCTCTGGGACAAAGACATGCACAGCTTTT 1507  
 Qy 468 GluPheArgValLysGluSerSerIleIleAlaProAlaProAlaGluAspValAspThr 487  
 Db 1508 CAGTTTGGACGTCAAAGAA---TCACTTATCTTAAAGCTGCCAACAGATGACTCATGAAC 1564  
 Qy 488 ProProAlaGlyLysLysValLysLysIleArgLeuTrpAlaAlaHisCysArgLysIleGln 507  
 Db 1565 CCTGCACAGAAAGAAAGAAACACACAGGTGTGGGCGGCACTGCAGAGAAATTCAG 1624  
 Qy 508 LeuLysLysAspGlySerSerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArg 527  
 Db 1625 CTGAAGAAAGATACAAATTTTACACAGAGTATTACTTCAACCTCTGTGACCAACCCACAC 1684  
 Qy 528 GlnProCysAspSerSerCysProCysValIleAlaGlnAsnPheCysGlyLysPheCys 547  
 Db 1685 CGTCCTGTGACAGACATGCCCTCCGATCATGACCCAGAACTTTGTGAAGATTTGCG 1744  
 Qy 548 GlnCysSerSerGluCysGlnAsnArgPheProGlyCysArgCysLysAlaGlnCysAsn 567  
 Db 1745 CAGTGCACCCCAAGACTGCCAGATGCTTTCCTGTGTTCGCTGTAAACTCAGTGCAT 1804  
 Qy 568 ThrLysGlnCysProCysLysIleuAlaValArgGluCysAspProAspLysCysLeuThr 587  
 Db 1805 ACCAAGCAGATGTCATGCTACTTGGCACTTGTGAGTGTGACCTGACTTGTGCTTACC 1864  
 Qy 588 CysGlyAlaAlaAspHisIleTrpAspSerLysAsnValSerCysLysAsnCysSerIleGln 607  
 Db 1865 TGTGGGGCTCAGACACATGGGACTGTAAAGTGTGCTCTTGCAGAAACATGCAGCATCAG 1924  
 Qy 608 ArgGlySerLysLysHisIleuLeuLeuAlaProSerAspValAlaGlyTyrGlyLysPhe 627  
 Db 1925 CGTGGCCCTCAAAACACCTGCTGTGGCGCTTCCAGATGTGGCCGATGGGGCACCTTC 1984  
 Qy 628 IleLysAspProValGlnLysAsnGlnPheIleSerGluTyrCysGlyGluIleIleSer 647  
 Db 1985 ATCAAGAGAGTCTGTGACAGAGATGAATTCATTTCTGAATATTTGGTGAAGCTATCTCT 2044  
 Qy 648 GlnAspGluAlaAspArgArgGlyLysValTyrAspLysTyrMetCysSerPheLeuPhe 667  
 Db 2045 CAGGATGAGCTGATCGTGAAGGAGAGGTCTATGATTAATCATGTCCAGCTTCTCTTC 2104  
 Qy 668 AsnLysAsnAsnAspPheValValAlaPheAlaThrArgLysGlyAsnLysIleArgPheAla 687  
 Db 2105 AACCTCAACAATGATTTGTAGTGTCTACCCGAAAGGAAACAAATTCGCTTGTGA 2164

Qy 688 AsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 Db 2165 AACCATTCAGTGAACCCCACTGTATGCAAGT 2200

RESULT 6  
 AKI40694  
 LOCUS  
 DEFINITION  
 Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
 enriched library, clone:B930092M06 product:enhancer of zeste  
 homolog 1 (Drosophila), full insert sequence.

ACCESSION  
 AKI40694  
 VERSION  
 GI:74180279  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.

## REFERENCE

1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)

## AUTHORS

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)

## JOURNAL

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Komuro, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)

## JOURNAL

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Aizawa, T., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, W., Gaasterland, T., Gibbs, C., King, B., Kochiwa, H.,  
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,  
 Quackenbush, J., Schriml, L.M., Stebbins, P., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
 Baldarelli, R., Barab, G., Blake, J., Boffelli, D., Bojunga, N.,  
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Gassmann, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamita, M., Lee, N.H., Lyons, P.,  
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schonbach, C., Sessa, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
 Toyono, O., Wang, K.H., Weitz, C., Whitaker, C., Whiting, L.,  
 Wyshaw-Borja, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kontecki, S.  
 and Hayashizaki, Y.  
 RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409 (6821), 685-690 (2001)

## JOURNAL

5 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Oseto, N., Saito, R., Suzuki, H., Yamanka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.F., Bult, C.,

## JOURNAL

688 AsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 Db 2165 AACCATTCAGTGAACCCCACTGTATGCAAGT 2200

## JOURNAL

RESULT 6  
 AKI40694  
 LOCUS  
 DEFINITION  
 Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
 enriched library, clone:B930092M06 product:enhancer of zeste  
 homolog 1 (Drosophila), full insert sequence.

## JOURNAL

ACCESSION  
 AKI40694  
 VERSION  
 GI:74180279  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.

Hume, D.A., Quackenbush, J., Schiraldi, L.M., Kanapin, A., Matelda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bragici, V., Chochia, C., Corbani, L.E., Couzens, S., Della, B., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Gozlik, A., Gough, J., Grimmond, S., Guetlinich, S., Hirokawa, N., Jackson, I.J., Jarvis, B.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.J., Konegaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltale, L., Marchionni, L., McKenzie, L., Miki, H., Nagahima, T., Numa, K., Okido, T., Pavan, W.U., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sadelin, A., Schneider, C., Sempile, C.A., Secou, M., Shmada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlested, C., Wang, Y., Watanabe, Y., Wells, C., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Alizawa, K., Arakawa, T., Fukuda, S., Haru, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

CONSRPT  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)  
12466851

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodius, R., Shimokawa, K., Bajic, V.B., Bremner, S.B., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aldins, A., Allen, J.E., Ambesi-Impombato, A., Apweiler, R., Aurali, R.N., Bailey, T.L., Banal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., Di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgi-Hemming, P., Gingeras, T.R., Gojovic, T., Green, R.B., Gregor, S., Harbers, M., Hayashizaki, Y., Hensch, T.K., Hirokawa, N., Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jait, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, J., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan, Badi, M., Madera, M., Marchionni, L., Matelda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottacchi-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.U., Pavoni, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Roest, B., Ruan, Y., Salzberg, S.L., Sadelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Sempile, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, B., Sugita, K., Sultana, R., Takanaka, Y., Taki, K., Tanno, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Wei, K., Yamashita, H., Zdobych, R., Zhu, S., Zimmer, A., Hilde, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, Z.T., Brusic, V., Quackenbush, J., Wahlested, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanabe, A., Okamura-Ohno, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.

CONSRPT  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)  
16141072

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H.,

Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Reghbi, M.A., Sadelin, A., Chalk, A.M., Mottacchi-Tabar, S., Liang, Z., Lenhard, B. and Wahlested, C.

CONSRPT  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

RIKEN Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)  
16141073

8 (bases 1 to 4194)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp),  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

1. 4194  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM\_DB:B930092M06"  
/db\_xref="taxon:10090"  
/clone="B930092M06"  
/tissue\_type="cerebellum"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
119..2371  
/note="unlabeled protein product; enhancer of zeste homolog 1 (Drosophila) [MGD|G1:1097695 GB|NM\_007970, evidence: BLASTN, 99%, match=4156]  
putative"  
/codon\_start=1  
/protein\_id="BAE24446.1"  
/db\_xref="GI:74180280"  
/translation="MRMDIASPTSCITYWKKKXSEYRLQLRLQANMGAKAL  
YVNFVAVQKTOILNEKKLKVQPPQPKPVSGHPLKCTTESIFPGDSODML  
RSNIVAVLPIMVSPLOONFHEVETVLCNIPYMDVYEBETETLEILNNYDG  
VHGHEMIKPSGLVSDAFLVLEIDALNPNVDEBDEGNDSDGQDSKDLPTTRR  
KRAHIBKSSKQKPPNDITFSAIASMFPENYCPDPMKERYRELTREKSDNALPPOC  
TPNIDGNASVQREOSLHSHFTLPCRCRKYDCLFIPFATVPRVYRKKEKLEIES  
PCGNDCLLEGKAYVAMLNPSKCSGRRRRPVVASCSNASAMATKEDGD  
RDNDGASSESNRSCPTPKQSPAPAOCLVAPSPVATGASLTFVFG  
TYFNPFSIRLGTGKOVFGAVESILKLPTELNNPQKKRKRRLNANCR  
KIQLKDNNTQVYVNPQCHPRPDCSTPCIMTQNFCKRQFCQSPDCCNRPFCGNC

Alignment Scores:

Pred. No.:	1,92e-210	Length:	4194
Score:	2389.50	Matches:	451
Percent Similarity:	76.4%	Conservative:	93
Best local Similarity:	63.3%	Mismatches:	115
Query Match:	62.7%	Indels:	53
DB:	6	Gaps:	13

US-10-773-302-2\_COPY\_1\_699 (1-699) x AK140694 (1-4194)

15 TrrArgybaRgVallysSerGluTyrMeArGdeuArgInleuYsArgPheArg 34

Db 170 TGGAAAAAGAAAGTGAATCTGACTATATGCGCTTGACAGCTCAACCGCTCCACGCA 229  
Qy 35 AlaAspGluValIysSerMetPheSerSerAsnArgGlnIleuGluArgThyGlu 54  
Db 230 AATATGGAGACAAAGGCTCTGTATGTGGCAATTTTGCAGAAAGTTTCAAGAAAAACCA 289  
Qy 55 IleuAsnGlnGlnIleuArgGlnIleuArgGlnIleuArgGlnIleuArgGlnIleu 74  
Db 290 ATCTCATGATGAAGTGAAGAACTTCGTGCTGAGCTGTTCCAGCAATGAACCGCTG 349  
Qy 75 SerSerIleuArgGlnIleuArgGlnIleuArgGlnIleuArgGlnIleuArgGln 93  
Db 350 AGTGGGCAACCTTTCTGAAAAGTGAACATAGAGCATTTTCCAGGGTTTGACAGC 409  
Qy 94 GluValIleProIleuValIleuAsnIleuValIleuValIleuValIleuValIleu 113  
Db 410 CAGGATATGTTGATGCGGTCTCTGACACTGTTGCACTGTTCCATCATGATTTCTGG 469  
Qy 114 SerProIleuGlnIleuAsnIleuValIleuAspGluThrValIleuIleuAsnIleu 133  
Db 470 TCCCTCTCCAGCAAGATTTTCATGATGAGATGACAGGTTTGTGCAATATTCATAC 529  
Qy 134 MetGluAspGluValIleuAspGlnAspGlyThrPheIleuGluIleuIleuAsnIleu 153  
Db 530 ATGGGTGACGAGTGAAGAAAGATGACATTTTCATGAAAGCTGATCAATCAATCAT 589  
Qy 154 AspGlyValIleuValIleuAspArgGlu-----CysGlyPheIleuAsnAspGlu 169  
Db 590 GATGGCAAGTCCACGCTGAGAAAGATGATCCCTGATCTGTGCTGATCAAGGATCT 649  
Qy 170 IlePheValGluIleuValIleuValIleuValIleuValIleuValIleuValIleu 187  
Db 650 GTGTTTCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 709  
Qy 188 -----AspArgGly-----AspArgProGluGlu-----ArgGlu 197  
Db 710 AACGACCCCTCCGATGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 769  
Qy 198 GluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 217  
Db 770 CGAAGGCCCATGCTATGAGGCAACAA-----AAGATTCAG-----AACAG 817  
Qy 218 PheProSerAspGlyIleuPheGlnIleuValIleuSerSerMetPheProAspGlyIleu 237  
Db 818 TTTTCAATGACATGATCTTCAGCGCCATGCTGATGATGATGATGATGATGATGAT 877  
Qy 238 GluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 257  
Db 878 GACGACATGAAGAGGATATCGAGCTGACAGATGATGACAGCCCAATGCACTTCC 937  
Qy 258 ProGluCysThrProAsnIleuAspGlyProAsnIleuValIleuValIleuValIleu 277  
Db 938 CTTCACTGACACCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 997  
Qy 278 LeuIleuSerPheIleuValIleuValIleuValIleuValIleuValIleuValIleu 297  
Db 998 CTGACACTTTTCCACACCTTTTCTGCGCGCTGTTTAATATGATGATGATGATGAT 1057  
Qy 298 ProPheIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 317  
Db 1058 CTTTCCACGCGCAACCCCAATGATATATATATATATATATATATATATATATATAT 1117  
Qy 318 LysProCysGlyProGlnCysThrGlnIleuValIleuValIleuValIleuValIleu 337  
Db 1118 GATCGGTGAGACAGACTGCTCTTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171  
Qy 338 LeuThrIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 357  
Db 1172 -----ATGCTGACAAACCTCGTCCAAAGCTCTCGGCGCGCGCGGAGAG 1219  
Qy 358 LeuProAsnAsnSerSerArgProSerThrProThrIleuValIleu-----GluSerIleu 376

Db 1220 CACCCAGTGTCAAGTCTTCCTGCTCCAAATGATCAAGCTTCTGCTATGCTGAACATA 1279  
Qy 377 AspThrAspSerAspArgGlnIleuValIleuValIleuValIleuValIleuValIleu 396  
Db 1280 GAAGAGACAGTATGATGACACTGCG----- 1306  
Qy 397 GluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 416  
Db 1307 -----AATGCTGAGGCTCCAGTTTCTGAGAGGCTAATCTGCTGCTGAGAC 1354  
Qy 417 ProIleuMetIleuProAsn-----IleGlu-----ProPro 427  
Db 1355 CCAACGAAACAGAAAGCCAGCTCCAGCCCACTCACTGCTGCTGCTGCTGCTGCTGCT 1414  
Qy 428 GluAsnValIleuValIleuValIleuValIleuValIleuValIleuValIleu 447  
Db 1415 GAGCGGTGATGATGACCGGAGCCGAAAGATCTTTTCCGAGTCTTCCACGCGACCTAT 1474  
Qy 448 TyrAspAsnPheCysAlaIleuValIleuValIleuValIleuValIleuValIleu 467  
Db 1475 TTCAACAACTTCTGCTCAATATGACAGGCTTCTGAGCAAGACATGCAAGAGTCTT 1534  
Qy 468 GluPheArgValIleuValIleuValIleuValIleuValIleuValIleuValIleu 487  
Db 1535 CAGTTTTCAGTCAAGAA-----TCACTTATCTTAAGTGCACATGATGATGATGATGAT 1591  
Qy 488 ProProArgIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 507  
Db 1592 CTTGACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTCT 1651  
Qy 508 LeuIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 527  
Db 1652 CTGAAGAAAGATTAACATTTTCAACAGTATTAATTAACCAACCTGATGACACCCAGAC 1711  
Qy 528 GluProCysAspSerSerCysProCysValIleuValIleuValIleuValIleuValIleu 547  
Db 1712 CTTCCGCTGATCAAGCAATGCTCCCTGATGATGATGATGATGATGATGATGATGAT 1771  
Qy 548 GluIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 567  
Db 1772 CAGTGCAGCCAGACTGCGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1831  
Qy 568 ThrIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 587  
Db 1832 ACCAAGAAATGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1891  
Qy 588 CysGlyValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 607  
Db 1892 TGTGGGCTCTCAAGCACTGGGACTGTAAGTGTGTCTTGCAGAAATCTGACATCCAG 1951  
Qy 608 ArgGlySerIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 627  
Db 1952 CGTGGCTCAAAAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2011  
Qy 628 IleuValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 647  
Db 2012 ATCAAGAGCTGCTGCAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2071  
Qy 648 GluAspGlnIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 667  
Db 2072 CAGGATAGGCTGATCTCTGAGGAGAGCTGATGATTAATTAATTAATTAATTAATTAAT 2131  
Qy 668 AsnIleuAsnAspPheValIleuValIleuValIleuValIleuValIleuValIleu 687  
Db 2132 AACCTCAACATGATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2191  
Qy 688 AsnIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 699  
Db 2192 AACCATTCAGTGAACCCCACTGTTATGCAAAAGTG 2227

RESULT 7  
CR857825  
LOCUS CR857825 4643 bp mRNA linear HTC 12-NOV-2004

DEFINITION Pongo pygmaeus mRNA; cDNA DKFZp469N1815 (from clone DKFZp469N1815).  
ACCESSION CR857825  
VERSION CR857825.1 GI:55726634  
KEYWORDS HTC.  
SOURCE Pongo pygmaeus (orangutan)  
ORGANISM Pongo pygmaeus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pongo.  
REFERENCE 1 (bases 1 to 4643)  
Koehler, K., Beyer, A., Mewes, H. W., Well, B., Amid, C., Oeanger, A.,  
Fobbo, G., Han, M. and Wiemann, S.  
CONSRMT The German cDNA Consortium  
TITLE Direct Submission  
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by BMFZ (Biomedical Research Center) at the  
Heinrich-Heine-University, Duesseeldorf/Germany within the cDNA  
sequencing consortium of the German Genome Project. This clone  
(DKFZp469N1815) is available at the RZPD Deutsches  
Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneid=DKFZp469N1815  
Further information about the clone and the sequencing project is  
available at http://mips.gsf.de/projects/cdna/  
location/Qualifiers  
1..4643  
/organism="Pongo pygmaeus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9600"  
/clone="DKFZp469N1815"  
/cfeature\_type="Kdney"  
/clone\_lib="469 (synonym: pkid1). Vector pSport1\_Sfi; host  
DH10B; sites Sfi1A + Sfi1B"  
/dev\_stage="adult"  
/note="enhancer of zeete homolog 1 (Homo sapiens)"  
1..4643  
/gene="DKFZp469N1815"  
119..2362  
/gene="DKFZp469N1815"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAH90081.1"  
/db\_xref="GI:55726634"  
/translation="MEIDNPPTSKCTYWKRYKSEYRLNQLKLANMGAALYVA  
NPAVQKTOILNEEMKRLVQVQSMKPGHPLKCTIESIFPGASOMLMRL  
NTVLVPIITMSVPLQONPVBDETVLGNIPYMDREVKEBETPIELINNYDKVHG  
EBEATPGSVLISDAVFLVDALNYSBBERGANDTSDGQDSKEILPYTRKRRH  
ALBGNKSSKQIPNDMLFSALASMPENGVDDMKRYRLITKESDNLALPQCTN  
IDGPNKSVQROSLASHFHLFCRCFYDCLHPFALPVPYKKNKIKIEPCG  
TDCFLALGAEAYALHNPRSKSGRRRHRIVASCSNLSAVALTEKSDSDRT  
GNDMASSSEBANSRCQPTPKOKAPAPOLCVBEASPVEVTEGABEELPVEVGT  
NPFCSIALILGTCCKOYFOPAVYVESLILKLPTEILAMPQCKKRGHILMAHCKIO  
LAKDNSSITQVNYTPCDHPRCSTCTCTINTQNCERFCQCNPDQNRPFQCRKTO  
CNTQKPCYTLVRECDPDLCTGASERMDCKVSCNCSIQRLKRLILAPSDVAG  
WGTFIKESVQKNEPISYCGELISQDEADRGKVDKTMSEFLFNALNDPVADATRK  
NKIRFANSVNPNCTAKYVMVNGDHRIGI FAKRAIQAGEELFLDYRYSGADALKVGI  
EREVDVL"

ORIGIN

Alignment Scores:

Pred. No.:	1,246-209	Length:	4643
Score:	2381.50	Matches:	448
Percent Similarity:	75.4%	Conservative:	89
Best Local Similarity:	62.9%	Mismatches:	122
Query Match:	62.5%	Indels:	53
DB:	6	Gaps:	10

US-10-773-302-2\_COPY\_1\_699 (1-699) x CR857825 (1-4643)

QY 15 TrrpArglybArgValLySerGluTyrMetArgLeuAArgGlnLeuLybArgPheArgArg 34

DB 161 TGGAAAGAAAAAGTAAATCTGAATACATGCACTTCGACAACTTAAAGCGTTTCAGGCA 220  
QY 35 ALaepGluValLySerMetPheSerSerArgbArgGlnLybLeuGluArgThrglu 54  
DB 221 AATATGGGTGCMAAGGCTTGTATGTGCAATTTTGCMAAGGTTCAAGAAAAACCCAG 280  
QY 55 ILeuAaGlnGlnTrrpLybGlnbArgArgGlnGlnProlValbLeuThreVal 74  
DB 281 ATCTCAATGAAAGAAATGAAAGGCTTGTGTCCAACTGTTCAGTCAATGAAGCTGTG 340  
QY 75 SerSerLeuArgLybThrArgGluCySerValThrSerArgPleu---AspPheProthr 93  
DB 341 TGTGACACCCCTTTCTCAAAAAGGTACACATAGACACATTTCCCGGATTTCCAGC 400  
QY 94 GlnValIleProlLeuLybThrLeuAaValaValaSerValProlMetYSerTrp 113  
DB 401 CAACATATGTTAATAGAGGTCACTGMAACAGTGTGATTTCCCATATGATTTCTGG 460  
QY 114 SerProlLeuGlnGlnbArgPheMetValGlnbArgGlnTrrpValbLeuHlaAaMlleProTyr 133  
DB 461 TCCCTCTTCCAAAGAACCTTATATGTAGAAAGATGACGGTTTGTGCAATATTCCTAC 520  
QY 134 MetGlyAepGluValLeuAaArgGlnbArgLybThrPheIleGlnGlnLeuIleYsAntYr 153  
DB 521 ATGGAGATGAAATGAAAGAAAGAAATGACCTTTATTTGAGAGCTATCAATACAT 580  
QY 154 AepGlyLybValbArgLybArgLyb-----CybGlyPheIleAaAepGlu 169  
DB 581 GATGGAAAGTCCATGTGTGAAGAAAGATGATCCCTGATCCGTCTGATTTAGTATGCT 640  
QY 170 IlePheValGlnLeuValAaAaMlleuGlnTrrpAaAaPasp----- 185  
DB 641 GTTTTTCGAGTTGTGTGATGCCCTGATACATCAATGAGAGAGAGAGGAC 700  
QY 186 -----AspAaPaspArgLybAaPaspProGluGlnArgGlu 197  
DB 701 AATGACACCTGATGAAAGAGATGACAGAGATGACAGAAAGATCTCCAGTAAAGAAAG 760  
QY 198 GlnLybGlnLybAaPleuGlnbArgbArgbArgLybGlnSerArgProProArgLyb 217  
DB 761 AGAAAGGACATGATGATGAAGGACAA-----AAGAGTTCAG-----AAACG 808  
QY 218 PheProSerArgLybIlePheGlnAaIleSerSerMetPheProArgLybGlyThrAla 237  
DB 809 TTCCCAATGACATGATCTTCAAGTCCCTCAATGCTCCCAAGTCCCTGAAATGCTGCCA 868  
QY 238 GlnGlnLeuLybGlnLybLybGlnLeuThrglnGlnGlnLeuProGlyValaLeuPro 257  
DB 869 GATGACATGAAAGAGAGATGACAGAGCTTAAAGAGATGACAGCCCAATGCACTTCC 928  
QY 258 ProGlnCybThrProAaMlleAepLybProAaMlleYsSerValGlnArgGlnbSer 277  
DB 929 CCGAGTGCACACCAACATGACAGGCGCCCAATGCAAGTCTGAGCAGCGGAGCAATCT 988  
QY 278 LeuHlaSerPheHlaThrLeuPheCybArgArgCybPheLybTrrpCybPheLeuHla 297  
DB 989 CTGCACTCTCTTCCACACACTTTTGTGCGGCGCTTAAATGCACTGCTTCCCTTAC 1048  
QY 298 PhePheHlaAaThrProAaMlleThrTrrpLybArgLybAaMlleThrAlaLeuAaPasp 317  
DB 1049 CTTTTCATGCAACCCCTTAATGATATTAACCAAGAAATTAAGAAATGACAGATGAAAC 1108  
QY 318 LybPProCybGlyProGlnCybTrrpMlleuGlnGlnLybAaLybGlnPheAaIlaAa 337  
DB 1109 GAACCATGTGCAACAGCTGCTTCTTGTGAGAGAGCAAGAGATAGCC----- 1162  
QY 338 LeuThralaGlnGlnIleLybThrProProLybArgProGlyGlybArgArgGlybArg 357  
DB 1163 -----ATGCTCCATTAACCCCGCTCAAGGCTGCTGCTGCTGCGGAGAGAGG 1210  
QY 358 LeuProAaMleAaSerArgProSerThrProThrIleAa---ValLeuGlnSerLyb 376

Db	1211	CACCAATAGTCACTGCTTCTGCTCCAAATGCTCAAGTTCTGCTGCTGAGACTTAA	1270
Qy	377	AspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyGluAsnAsnAspGlyGlu	396
Db	1271	GAAGAGACAGTGCACGAGACACAGC-----	1297
Qy	397	GluGluGluGlyGlyAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThr	416
Db	1298	-----AATGACTGGGCGCTCCAGTTCTTCAGAGGCTAACTCGCTGCAGACT	1345
Qy	417	ProTlelysmelysProAsnIleGluProPro-----	427
Db	1346	CCCCCAAAACAGAAAGGCTAGTCAGAGCCCACTCACTGCTGCTGAGGAAACACCTCG	1405
Qy	428	GluAsnValGluTyrSerGlyValAlaGluAlaSerMetPheArgValIleGlyThrTyr	447
Db	1406	GAGCTGTGAAATGACATCGGGGCTGAAAGATCTCTTTTCCAGTCTTCCAGTGCACCTAC	1465
Qy	448	TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThrTyrThrCysArgGlnValTyr	467
Db	1466	TTCAACAACCTCTGTTCATATAGCCAGGCTTCTGGGGAACCAAGACGTGCACAGGTCTTT	1525
Qy	468	GluPheArgValIleGlySerSerIleIleAlaProAlaProAlaGluAspValAspThr	487
Db	1526	CAGTTTGCAGTCAAGAA--TCACTTATCTTGAAGCTGCCAAGGATGACTCATGAAC	1582
Qy	488	ProProArgIleGlyGlyValArgIleAsnArgPheProGlyCysArgCysValIleGln	507
Db	1583	CCCTCAGAGAGAAAGAAAGAAAGACAGATCTGGGTGGTGCACCTGAGAGAAAGTTAG	1642
Qy	508	LeuIleGlyAspArgIleSerSerAsnIleValIleAsnTyrGlnProCysAspHisProArg	527
Db	1643	CTGAAGAAAGATTACTTTCCACACAACTGTACAACTCAACCTGCGACACCCCAAC	1702
Qy	528	GlnProCysAspSerSerCysProCysValIleAlaGlnAsnPheCysGluIlePheCys	547
Db	1703	CGCCCTGTGACAGACCTGCCCTGCATCATGACTCAGAAATTTCTGAGAAAGTTGTC	1762
Qy	548	GlnCysSerSerGluCysGlnAsnArgPheProGlyCysArgCysValAlaGlnCysAsn	567
Db	1763	CAGTGCACCAAGACTGTCAATATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1822
Qy	568	ThrIleGlnCysProCysTyrIleAlaValAlaArgGluCysAspProAspLeuCysLeuThr	587
Db	1823	ACCAAGCAATGTCTTGTCTATCTGCGACGTGCAGATGTGACCCCTGACCTGTGCTCAC	1882
Qy	588	CysGlyAlaIleAsnIleTyrAspSerIleAsnValSerCysIleAsnCysSerIleGln	607
Db	1883	TGTGGGCTTCAAGACCTGGACCTGCAGAGGTGTTCTCTGTAATAACTGCGACATCCAG	1942
Qy	608	ArgGlySerIleGlyAsnIleLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIlePhe	627
Db	1943	CGTGGCTTAAAGACACCTGCTGCTGCCCTCCCTCAATGTGGCCGATGGGGACCTTC	2002
Qy	628	IleIleAspProValGlnIleAsnGlnPheIleSerGluTyrCysGlyGluIleIleSer	647
Db	2003	ATTAAGAGAGCTGTGCAGAGACCAATTCATTTCTGATATCTGTGGAGACTCATCTCT	2062
Qy	648	GlnAspGluAlaAspArgArgGlyIleValIleTyrAspIleTyrMetCysSerPheIlePhe	667
Db	2063	CAGAGTAGGCTGATGCAGCGGAAAGGCTATGACAAATCATTCACAGCTCTCTCTTC	2122
Qy	668	AsnLeuAsnAsnAspPheValIleValAspAlaThrArgIleGlyAsnIleIleArgPheAla	687
Db	2123	AACTTAATTAATGATTTTGTATGTGATCTACTCGAAAGGAAACAAATTCGATTTC	2182
Qy	688	AsnHisSerValAsnProAsnCysTyrAlaIleValIleValIleValIleValIleVal	699
Db	2183	AATCATTCAGTGAATCCCAAGCTGTTATGCCAAAGTG	2218
RESULT 8			
AKI38942			
LOCUS	AKI38942	2142 bp	mRNA linear HTC 21-SEP-2005

DEFINITION	Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530023C07 product:enhancer of zeste homolog 1 (Drosophila), full insert sequence.
ACCESSION	AKI38942
VERSION	AKI38942.1 GI:74223861
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
TITLE	2
JOURNAL	3
PUBMED	11042159
REFERENCE	4
AUTHORS	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arai, K., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, K., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Pleschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pecole, G., Quackenbush, J., Schriml, L.M., Scudlitz, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzatelli, U., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilmink, L., Wyszewski, B., Yoshida, K., Haegawa, Y., Kawai, J., Kohetsuki, S. and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	11076861
REFERENCE	5
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oshio, K., Saito, R., Suzuki, H., Yamana, K., Kiyosawa, H., Yagi, K., Tomaru, Y., Haegawa, Y., Nogami, A., Kiyosawa, H., Gojobori, T., Baldarelli, R., Hill, D., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanpin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Bruns, V., Chocho, C., Corbett, L.E., Cousens, S., Dalla, E., Dragan, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, J.D., Jarvis, B.D., Kanai, A., Kawai, J., Kawasawa, Y., Kedzierski, R.M., King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
CONSRMT	
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
PUBMED	11217851
REFERENCE	
AUTHORS	







Db 266 ATCTCATGTAAGAGTGGAGAACTTCGTGTCAGCCTGTTCAAGCCATACCCGCTG 325  
Qy 75 SerSerLeuAArgLYThrArgGLuCySerValThrSerAAspLeu---AspPheProThr 93  
Db 326 AGTGGGCAACCTTTTCGAAAAGGTACATAGAGACATTTTCCAGGGTTTGACAGC 385  
Qy 94 GlnValIleProLeuYsThrLeuAsnAlaValAspValProIleMetYsSerTrp 113  
Db 386 CAGATATATGTTGATGCGGCTCTCTGAACACTGTGACAGTGTCCATCATATATCTCGG 445  
Qy 114 SerProLeuGlnGlnAsnPhMetValGluAAspGluThrValIleuHisAsnIleProTrp 133  
Db 446 TCCCTCTCCAGCAATTTTCATGTGTGAGATGAGCGGTTTGTGTCAATATTCATAC 505  
Qy 134 MetGlyAAspGluValIleuAAspGluAAspGlyThrPheIleGluGluIleuIleGlyAsnTrp 153  
Db 506 ATGGGTGACGAGGTGAAGAGAAAGATGACATTTCATGAGAGCTGATCATATACTAT 565  
Qy 154 AAspGlyValHisGlyAAspArgGlu-----CySgLYPheIleAsnAAspGlu 169  
Db 566 GATGGCAAAAGTCCACGCTGAGAGAGAGATGATCCCTGATCTGTGCTGATCAGCGATCT 625  
Qy 170 IlePheValGluLeuValAsnAlaLeuGlyGlnTrpAsnAAspAAspAAsp--- 187  
Db 626 GTGTTTCTGAGCTGTGTGATGCTCCTCAACCACTACTCTGATGAGAGAGAGAGAGAG 685  
Qy 188 -----AAspAAspGly-----AAspAAspProGluGlu-----ArgGlu 197  
Db 686 AACGACCCCTCCGAGGAAAGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 745  
Qy 198 GlnLYsGlnLYsAAspLeuGluAAspHisArgAAspAAspLYsGluSerArgProArgLYs 217  
Db 746 CGAAGAGCGCATGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 793  
Qy 218 PheProSerAAspLYsIlePheGluAlaIleSerSerMetPheProAAspLYsGlyThrAla 237  
Db 794 TTTCCAAATGACATGATCTTCAGCGCCATGCGCTGATGTTCCGAGAAATGGTCTCCT 853  
Qy 238 GlnGluLeuLYsGluLYsTrpLYsGluLeuThrGlnGlnIleuProGlyValIleuPro 257  
Db 854 GACGACATGAG 913  
Qy 258 ProGluCySerThrProAsnIleAAspGlyProAsnAlaLYsSerValGlnArgGluGlnSer 277  
Db 914 CCTCAGTGCACACCCACATGATGATGCGCCCAACGCAAGTCACTGACAGGAGAGAGCT 973  
Qy 278 LeuHisSerPheHisThrLeuPheCyArgArgLYsPheLYsTrpAAspCyPheLeuHis 297  
Db 974 CTGCACTCTTTCCACACCTTTTCTGCGCGCGCTGTTTAAATATGACTGCTCTTCCAC 1033  
Qy 298 ProPheHisAlaThrProAsnThrTrpLYsArgLYsAsnThrGlnThrAlaLeuAAspAsn 317  
Db 1034 CTTTTCACGCGCCACCCAAATGATATATAGCCAGAACAGAGAAATCAAGATTAGCCA 1093  
Qy 318 LYsProCYsGlyProGlnCYsTrpGlnHisLeuGlnGluAlaLYsGluPheAlaIleAla 337  
Db 1094 GAACCGTGTGAGACAGACTGCTCTTTTGTGCTGAGAGAGAGAGAGAGAGAGAGAGCC 1147  
Qy 338 LeuThralaGluArgIleLYsThrProProLYsArgProGlyLYsArgArgArgLYsArg 357  
Db 1148 -----ATCTGCACAAACCTCGTCCAAAGTCTCGGCGCGCGCGCGAGAGG 1195  
Qy 358 LeuProAsnAAspSerArgProSerThrProThrIleAsnValIleu---GluSerLYs 376  
Db 1196 CACCCAGTGTGATAGGCTCTCTGCTCAATGATCATAGCTTCTGCTATGCTGAAACTTAA 1255  
Qy 377 AAspThrAAspSerAAspArgGluAlaGlyThrArgLYsGlyGluAsnAAspLYsGlu 396  
Db 1256 GAAAGAGACAGTATGAGACACTGCG----- 1282  
Qy 397 GlnGlnGlnLYsLYsAAspGluThrSerSerSerSerSerGluAlaAsnSerArgCYsGlnThr 416

Db 1283 -----AATGACTGGGCTCCATGTTCTTCAAGAGCTAATCTCCCTGTGACAGC 1330  
Qy 417 ProIleLYsMetLYsProAsn-----IleGlu---ProPro 427  
Db 1331 CCCACAGAAACAGAAAGCAGATCCAGCCGACGCTCAGCTGTGTGTGTGGAAAGCCCTCA 1390  
Qy 428 GlnAsnValGluTrpSerArgLYsAlaGluAlaSerMetPheArgValIleuIleGlyThrTrp 447  
Db 1391 GAGCGGTGAGAAAGAGCCGAGCGGAAGATCTCTTTCCAGGCTTCCACAGGCACTAT 1450  
Qy 448 TyrAAspAAspPheCyAlaIleAlaArgLeuIleGlyThrYsThrCYsArgGlnValTrp 467  
Db 1451 TTCAACAATCTTCTCTCAATAGCCAGGCTTCTGGGAGCAAAAGCATGCAAGAGGCTTT 1510  
Qy 468 GlnPheArgValLYsGluSerSerIleIleAlaProAlaProAlaGluAAspValAAspThr 487  
Db 1511 CAGTTTCAGTCAAGAA---TCACTTATCTTAAGCTGCCAACAGATGAGCTCATGAC 1567  
Qy 488 ProProArgLYsLYsLYsValArgLYsHisArgLeuThralaAlaHisCYsArgLYsIleGln 507  
Db 1568 CTTGCACAGAAAGAGAAAGAAACACAGAGTTGTGGCCGACACCTGCAGAGAAATTCAG 1627  
Qy 508 LeuLYsLYsAAspGlySerSerSerAsnHisValTrpAsnTrpGlnProCYsAAspHisProArg 527  
Db 1628 CTGAAGAAAGATTAACAATTTCTACAGAGTATATACCAACCTCTGTACACCCAGAC 1687  
Qy 528 GlnProCYsAAspSerSerCYsProCYsValIleAlaGlnAsnPhCYsGluLYsPheCYs 547  
Db 1688 CGTCCGTGTACAGACATGACCCCTGCATCATGACCCAGAACTTTGTGAAAAGTTCTGC 1747  
Qy 548 GlnCYsSerSerGluCYsGlnAsnArgPheProGlyCYsArgCYsLYsAlaGlnCYsAsn 567  
Db 1748 CAGTGCAGCCACAGCTCCAGAAATCGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1807  
Qy 568 ThrLYsGlnCYsProCYsTrpLYsLeuAlaValArgGluCYsAAspProAAspLeuCYsLeuThr 587  
Db 1808 ACCAAGCAATGTCTCATCTACTTGTGAGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1867  
Qy 588 CYsGlyAlaAlaAspHisTrpAAspSerLYsAsnValSerCYsLYsAsnCYsSerIleGln 607  
Db 1868 TGTGGGCTCAGAGACTGGAGCTGTAAAGGTGTGCTTGCAGAAAACCTGACAGATCCAG 1927  
Qy 608 ArgGlySerLYsLYsHisLeuLeuLeuAlaProSerAAspValAlaGlyTrpGlyIlePhe 627  
Db 1928 CGTGGCTCAAAAAGCACTGCTGTGCTGCTCCATGTGTGCGAGATGGGAGCACTTC 1987  
Qy 628 IleYsAAspProValGlnLYsAsnGluPheIleSerGlnTrpCYsGlyGluIleIleSer 647  
Db 1988 ATCAAGAGTGTGTGCAAGAGAAATGAATTCATTTCTGAATATGTGTGAGCTCATCT 2047  
Qy 648 GlnAAspGluAlaAAspArgArgLYsValValTrpAAspLYsTrpMetCYsSerPheLeuPhe 667  
Db 2048 CAGATGAGGCTGATCTGTCAGAGAGAGAGATGATATTAATATCATGTCCAGTCTCTTC 2107  
Qy 668 AsnLeuAsnAAspPheValValAAspAlaThr 678  
Db 2108 AACCTCAACATGATTTGTATGATGATGCTAAC 2140  
RESULT 9  
AY402398 2215 bp DNA linear GSS 15-DEC-2003  
LOCUS Mue musculus E2H1 gene, VIRIDAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY402398  
VERSION AY402398.1 GI:39758384  
KEYWORDS GSS.  
SOURCE Mue musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2215)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejalraival,A.,



```
Db      1513  TTCTACACAAAGTATATACTACCAACCTTGACCAACCAACCGTGTGACAGAC 1572
Qy      533    rCyProCyValIleaglAanpheCyGluYpHeCyGlnCySerSerGluCy 553
Db      1573  ATGCCCCCTGCATCATGACCCGAACTTTTGGAAAAGTTCTGCCAGTCACCAAGACTG 1632
Qy      553  AGInaAnpRheProGlyCyArgCyValaglInCySAnThrxYpGlnCyProCy 573
Db      1633  CCAGATCGCTTCTCGGTGTGCTGTAAGACTCAGTCAATACCAAGCAATGTCAATG 1692
Qy      573  GTTLeuAlaValArgGluCyAapProAapLeuCyEleuThrxYpGluYalaAapR 593
Db      1693  CTACTTGCAGCTTGTGTAGTGTGACCTGACTTGTGCTCACTGTGGGCTTCAAGCA 1752
Qy      593  GTTApSerIySAnValSerCyValaAnCySerIleaglInArgIySerIySAnH 613
Db      1753  CTGGGACTGTAAAGGTGTGTCTTGCMAAACTGCAGCATCCAGCGTGCCTCAAAAACA 1812
Qy      613  bLeuLeuAlaProSerAapValAlaglYTrpGlyTLePheIleYpAapProValG 633
Db      1813  CCTGCTGTGGCCCTTCCGATGTGGCCGATGGGCGACCTTCATCAAGAGTCTGTGCA 1872
Qy      633  nLyAaAngIuPheIleSerGluYTrCyGluGluIleIleSerGlnAapGluAlaAapR 653
Db      1873  GAAGAATGAATTCATTTCTGAATATGTGTGAGCTCATCTCTCAGATGAGGCTGATCG 1932
Qy      653  GAAGIYLySValYTrAapIyTrMetCySerPheUapPheAnLeuAaAnAapRph 673
Db      1933  TCGAGGAGAGGTCTATGATTAATACATCTCCAGCTTCTTCAACCTCAACATGATTT 1992
Qy      673  eValValaAapAlaThrxArgIyGluYpAAnLySileArgPheAlaAnHleSerValAapR 693
Db      1993  TGTATGTGATCTCTACCCGAAAGGAAACAATTCGCTTCAAAACCTCAGTGAAGCC 2052
Qy      693  cAaAnCyTrAlaYpVal 699
Db      2053  CAACTGTATGCCAAAGTG 2071
```

RESULT 10  
AY402396 2244 bp DNA linear GSS 15-DEC-2003  
LOCUS AY402396 Homo sapiens EZH1 gene, VIRIDAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY402396  
VERSION AY402396.1 GI:39758382  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 2244)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tanenbaum,D.M., Clavello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA

TITLE This sequence was made by sequencing genomic exons and ordering  
COMMENT them based on alignment  
FEATURES Location/Qualifiers  
SOURCE 1..2244

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..2244  
/gene="EZH1"  
/locus\_tag="HOM1210"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,91e-197 Length: 2244  
Score: 2243.50 Matches: 430  
Percent Similarity: 70.8% Conservative: 74  
Best Local Similarity: 60.4% Mismatches: 155  
Query Match: 58.9% Indels: 53  
DB: 14 Gaps: 10

US-10-773-302-2\_copy\_1\_699 (1-699) x AY402396 (1-2244)

```
Qy      15  TTPrArgLySArGValIySersGluYTrMetArgLeuArgGluInLeuYpArgPheArg--- 33
Db      43  TGGAAAAGAAAGTGAATCTGAATACATGCACTTGCACACTTAACAGCTTCAAGCA 102
Qy      34  ArgAlaAapGluValIySersMetPheSerSerAanArgGluYsileuGluArgThr 53
Db      103  AATATGGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 162
Qy      54  GluIleuAaAngIuTrpIyGlnArgArgIleGlnProValHleIleuThrSer 73
Db      163  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 222
Qy      74  ValSerSerIeUarGlyThrxArgGluCySerValThrSerAapLeuAapPheProThr 93
Db      223  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 282
Qy      94  GluValIleProLeuYsThrxLeuAanIleValAlaSerValProIleMetYTrSerTrp 113
Db      283  CAACATATGTTAATAGAGTCACTGAACAGCTTGCAATTTGTTCCCATCATGATTCCTGG 342
Qy      114  SerProLeuGlnGlnAanPheMetValGluAapGluThrxValIleuHleAanIleProYTr 133
Db      343  TCCCTCTCCAAAGCAACTTATATGTGTGAAGATGACAGGTTTGTGCAATATTCCTTAC 402
Qy      134  MetGluYpArgIuValIleAapGlnAapGlyThrxPheIleGluGluIleuYsAnThrx 153
Db      403  ATGGAGATGAAAGTGAAGAAAGAGATGACATTTTATGAGAGCTGATCATATTAATCTAT 462
Qy      154  AapGlyYpValHleGlyAapArgGlu-----CyGlyPheIleAanAapGlu 169
Db      463  GATGGAAAGTCCATGTGTGAAGAGATGATCCCTGATCCGTTGATTAATGATGATGCT 522
Qy      170  IlePheValGluLeuValaAnIleuGlyGlnIYTrAanAapAapR----- 185
Db      523  GTTTTCTGAGTTGTGCGATGCCCTGAATCAGTACTCAATGAGAGAGAGAGAGGACAC 582
Qy      186  -----AapAapAapAapGlyYpAapProGluGluArgGlu 197
Db      583  AATGACACTCAGATGAAAGCAGATGACGCAAGAAAGATCTGCCAGTAAACAAGAG 642
Qy      198  GluYpGlnYpAapLeuGluAapHleArgAapAapIySersArgTrpProArgLyS 217
Db      643  AGAAAGGACATGCTATTGAAGGCACAA-----AAGAGTTCAAG-----AAACAG 690
Qy      218  PheProSerAapIySilePheGluAlaIleSerSerMetPheProAapIyGlyThrxAla 237
Db      691  TTCCCAATATGACATGATCTTCAAGTCAATGCTCATATGTTCCCTGAGATGTGTGCCA 750
Qy      238  GluGluLeuYpGluYsTrYpGluLeuThrxGluGlnGlnLeuProGlyAlaAapPro 257
Db      751  GATGACATGAAAGGAGATGATGCACTAACAAGATGTCAAGCCCAATGCACTTCCC 810
Qy      258  ProGluCyThrProAanIleAapGlyProAanAlaYpSerValGlnArgGluGlnSer 277
Db      811  CTTGAGTGCACCAACATCATGATGATGCCCAATGCCAAGTCTGTGACGGGAGCAATCT 870
```

```

QY 278 LeuHisSerPheHisThrLeuPheCysArgArgCysPheLeuTyrAspCysPheLeuHis 297
DB 871 CTGACCTCTTCCACACACTTTTGGCGGCTCTTAATAAGACAGCTCTCTTCCAC 930
QY 298 ProPheHisAlaThrProAsnThrTyrIlyAsArgIyAsenThrGluThrAlaLeuAspAsn 317
DB 991 CTTTTCATGCGCACCCCTTAATGTATTAACCCAGAAATTAAGAAATCAAGTTGACCA 990
QY 318 LysPProCysGlyProGlnCysTyrGlnHisLeuGluGlyAlaLeuGluPheAlaAla 337
DB 991 GAACCATGTGGCACAGACTGCTCTCTTTGCTGGAAGGACAAAGAAATGATCC----- 1044
QY 338 LeuThrAlaGluArgIleLeuThrProProlyAsArgProGlyIlyArgArgArgIly 357
DB 1045 -----ATGCTCCACACCCCGCTCCAGTGTCTGTGTGCGCGGAGAAAG 1092
QY 358 LeuProAsnAsnSerSerArgProSerThrProThrIleAsn---ValLeuGluSerIys 376
DB 1093 CACCAATAGTCAGTGTCTTCCGTCCATGCTCAGCTCTGTGTGCTGAGACTAA 1152
QY 377 AspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyIlyAsnAsnAspIlyGlu 396
DB 1153 GAAGGAGACAGTGAAGGAGACAGGC----- 1179
QY 397 GluGluGluIlyIlyAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThr 416
DB 1180 -----AATGACTGGGCGCTCCAGTTCTTCAAGGCTTAACCTCTGCTGACACT 1227
QY 417 ProIleYsMetIyProAsnIleGluProPro----- 427
DB 1228 CCCACAAACAGAAAGGCTAGTCCAGCCCACTCACTCTGCTGAGTGAAGACCTCG 1287
QY 428 GluAsnValGluTyrSerGlyAlaGluAlaSerMetPheArgValIleuIleGlyThrTyr 447
DB 1288 GAGCTGTGATGATGAGCTGGGCTGAAGATCTTTTTCAGTCTTCCATGCGACCTAC 1347
QY 448 TyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThrIlyThrCysArgGlnValTyr 467
DB 1348 TTCAACCACTCTGTTCATATGACCGGCTCTTGGGAGACCAAGACCTGACAGAGGCTCTT 1407
QY 468 GluPheArgValIyGlySerSerIleIleAlaProAlaProAlaGluAspValAspThr 487
DB 1408 CAGTTTGCAGTCAAGAA---TCACTTATCTGAAAGCTGCCAAGATGAGCTCATGAAAC 1464
QY 488 ProProArgIlyIlyAsArgIyHisIaArgLeuTyrAlaAlaHisCysArgIyIleGln 507
DB 1465 CCTCACAAGAAAGAAAGAAAGAACAGCATGTTGGGCTGACACTGACAGGAAGATTGAG 1524
QY 508 LeuIlyIlyAspGlySerSerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArg 527
DB 1525 CTGAAGAAAGATTAACCTTCCACACAGATGATCACTCAACCTCTGACCAACCCCAAC 1584
QY 528 GlnProCysAspSerSerCysProCysValIleAlaGlnAsnPheCysGluIlyAspHis 547
DB 1585 CGCCCTGTGACAGCACTGCGCCCTGACATCAAGTCAAGATTTCTGTGAAGAGTTGCGC 1644
QY 548 GlnCysSerSerGluCysGlnAsnArgPheProGlyIlyAsArgIyAlaGlnCysAsn 567
DB 1645 CAGTGCAACCAAGACTGTCAAGATGTTTCCCTGTGCTGTGCTGTGAACCAAGGCAAT 1704
QY 568 ThrIlyGlnCysProCysTyrIleuAlaValArgGluCysAspProAspLeuCysLeuThr 587
DB 1705 ACCAAGCATGTCTTGTCTATCTGACAGTGGAGATGTGACCCCTGACCTGTGTCTCAC 1764
QY 588 CysGlyAlaAlaAspHisIleTyrAspSerIyAsnValSerCysIyAsnCysSerIleGln 607
DB 1765 TGTGGGCTCTCAAGACTGGGACTGCAAGGTTTCTGTAAATCTGACAGATCCAG 1824
QY 608 ArgIySerIlyIlyHisIleuIlyLeuAlaProSerAspValAlaGlyTyrGlyIlyPhe 627
DB 1825 CGTGGACTTAAGAAAGCACTGTGCTGCGCCCTCTGATGTGGCGGAGTGGGACCTTC 1884

```

```

QY 628 IleYsAspProValGlnIlyAsnGluPheIleSerGlnTyrCysGlyGluIleIleSer 647
DB 1885 ATTAAGAGCTCTGTGCAGAAAGACAAATTCATTCTGAATCTGTGGAGCTCATCTCT 1944
QY 648 GlnAspGluAlaAspArgArgGlyIlyValIlyAspIlyTyrMetCysSerPheLeuPhe 667
DB 1945 CAGATGAGGCTGTATCGACCGGGAAGGCTATATGCAATATCATGTCCAGCTTCTCTTC 2004
QY 668 AsnLeuAsnAsnAspPheValIlyAspAlaThrArgIyGlyAsnIlyIleArgPheAla 687
DB 2005 AACCTCATATATGATTTTGTGTGTGATGCTCTCTCGAAAGAAACAAATTCGATTGCA 2064
QY 688 AsnHisSerValAsnProAsnCysTyrAlaIlyVal 699
DB 2065 AATCATTCATGATATCCCACTGTTATGCCAAAGTG 2100

RESULT 11
AY402397 2183 bp DNA linear GSS 15-DEC-2003
LOCUS AY402397
DEFINITION Pan troglodytes BZH1 gene, VIRUTAL TRANSCRIPT, partial sequence.
ACCESSION AY402397
KEYWORDS AY402397.1 GI:39758383
SOURCE GSS.
ORGANISM Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
1 (bases 1 to 2183)
REFERENCE Clark,A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tenenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.C.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL PUBMED 14671302
REFERENCE 2 (bases 1 to 2183)
AUTHORS Clark,A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tenenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.C.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Source Location/Qualifiers
1..2183
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>2183
/gene="BZH1"
/locus_tag="HGM1210"

ORIGIN
Alignment Scores:
Pred. No.: 3,15e-189 Length: 2183
Score: 2157.00 Matches: 406
Percent Similarity: 75.8% Conservative: 65
Best Local Similarity: 65.4% Mismatches: 98
Query Match: 56.6% Indels: 52
DB: 14 Gaps: 9

US-10-773-302-2_COPY_1_699 (1-699) x AY402397 (1-2183)
QY 105 AlaservAlaProIleMetTyrSerTyrSerProLeuGlnGlnAsnPheMetValGluAsp 124
DB 316 GCATTGTTCCTCCATCATGATATTCCTGTGCTCCCTCTCAACAGAACTTTATGTGAGAGAT 375
QY 125 GluThrValIleuHisAsnIleProTyrMetGlyIlyAspGluValIleuAspGlnAspGlyThr 144

```

Db 376 GAGCGGTTTGTGCAATATTCCNNTACATGGGAGATGAAGTGAAGAAAGAAATGAGACT 435  
 145 PheileglugluLeuilelyeAanlyzAapgllylyeVal1H:sglyAapAaglu----- 162  
 436 TTTATTTAGAGGTTTGAATCAATTAATCAATGGAAGAGTCCATGGTGAAGAAAGATGATC 495  
 163 -----CysglyPhe1leAanpAglu1lePheVal1GluVal1Aanp1leAeu1Gln 180  
 496 CCTGGAATCCGTTCTGATTAAGTGAATGCTGTTTCTCGAGTGTGCGATGCCCTGAATCAG 555  
 181 TyranAanpAapAap-----AapAapAap 188  
 556 TACTCAGATGAGAGAGAGAGAGGCAATGCACTCAGATGGAAGACAGATGACACC 615  
 189 AapgllyAapAapProglugluAaggluglu1lylyeVal1lyeAe1leAapAap 208  
 616 AAAGAAATCTGCGACATGAACAAAGAAAGAAAGCAATGCTATTGAAGGCAACAA--- 672  
 209 AaplyygluSerAapProAaggluAaplyyAapPhePheSerAaplyy1lePheglu1leSer 228  
 673 ---AAGAGTTCAGAG-----AAACAGTTCACCAATGACATGATCTTCAGTGCATTTGCC 723  
 229 SerAapPheProAaplyygllyThral1aglu1lyeAagglu1lyylyylyyglu1leThr 248  
 724 TCAATGTTCCCTGGAATAGTGTGCCAGATGACATGAAGAGAGGTTCGAGAGCTAACA 783  
 249 Gluglu1leuProglu1leAeuProglu1lyyThyProAan1leAapgllyProAan 268  
 784 GAGATGTCAGACCCCAANNCACTTCCCTCAGTGCACACCCCAATGATGATGAGCCCAAT 843  
 269 Al1ylyeSerVal1Gln1agglu1leu1leu1leu1leu1leu1leu1leu1leu1leu 288  
 844 GCGAAGCTGTGCGACGCGAGCAATCTGCACTCTTCCACACATTTTGTGCGCGGC 903  
 289 CysPhe1lyyTyAapPhe1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 308  
 904 TGCCTTAATACGAGCTGCTCTCTNNNNCTTTTCAATGCGACCCCTAATGATTAACAGC 963  
 309 LyAanThrglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 328  
 964 AAGAAATTAAGAAATCAAGATTAAGAAATCAAGATTAAGAAATCAAGATTAAGAAATCA 1023  
 329 Gluglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 348  
 1024 GAAAGAGCAAGAGATGACC-----ATGCTCCCAACACCCCGCTCC 1065  
 349 ArgProglu1lyyAagglu1lyyArgleuProAanAanSerAapPheProSerThyPro 368  
 1066 AAGTGTCTGT 1125  
 369 Thr1leAan---Val1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 387  
 1126 TCAACCTCTGCTGT 1179  
 388 Thr1lyglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 407  
 1180 -----AATGATGCGGCTCCAGTCT 1200  
 408 Serglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 427  
 1201 TCAAGGCTAATCTCTGCTGT 1260  
 428 -----GluAanVal1Glu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 438  
 1261 CAACCTGCTGATGAGAGCAACCTCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
 439 MetPheArgVal1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 458  
 1321 CTTTTCGAGCTTCCATGCGACCTACTTCAACCACTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380  
 459 GllyThrylyyThyTyAagglu1lyyglu1leu1leu1leu1leu1leu1leu1leu1leu1leu 478

Db 1381 GGGACCAAGACGTGCAAGCAGCTCTTCACTTGTGCACTCAAGAA---TCACTTATCTGT 1437  
 479 ProAlaProAla1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 498  
 1438 AAGCTGCCAAGATGAGTGCATGAAACCTCTCAAGAAAGAAAGAAAGCAAGATTTG 1497  
 499 TTTTAA1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 518  
 1498 TGGCTGTGCAACGTGCAAGAAAGATTCAGCTTGAAGAAAGATTAATCTTCCACAAAGTTCAC 1557  
 519 AantTyglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 538  
 1558 AACTCAACCACTCTGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1617  
 539 Alaglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 558  
 1618 ACTCAAGAAATTTCTGTGAGAGTTCCTGCACTGCAACCAACCAACCAACCAACCAACCAAC 1677  
 559 GlyCyAagglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 578  
 1678 GGGCTGTGCTGTGAAGCCAGTGCATTAACCAACCAACCAACCAACCAACCAACCAACCAAC 1737  
 579 GlucyAapPheAapPhe1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 598  
 1738 GAATGTACCTCTGACCTGTGTCTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1797  
 599 ValSerCyAagglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 618  
 1798 GTTTCCTGTGAAGAAATGCAACATTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1857  
 619 SerAapVal1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 638  
 1858 TCGATGTCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1917  
 639 Serglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 658  
 1918 TCGAATACCTGT 1977  
 659 AaplyyTyAapPhe1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 678  
 1978 GACAAATACATGTGCGAGCTTCTCTTCAACCTCAATATGATTTGTGTGTGTGTGTGTGTGT 2037  
 679 Arglyglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu 698  
 2038 CGAAAGAAACAAATTCATTTGAAATTCATGTAATCCCACTGTATGCCAA 2097  
 699 Val 699  
 2098 GTG 2100  
 Db 2098 GTG 2100  
 RESULT 12  
 AKI53835  
 LOCUS  
 DEFINITION  
 full-length enriched library, clone: C920016A05 product: enhancer of zeste homolog 2 (Drosophila), full insert sequence.  
 AKI53835  
 VERSION  
 AKI53835.1 GI:74150295  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 2  
 10349636  
 Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtractions of cap-trapper-selected cDNAs to

Prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Ktsunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okada, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multiplexed sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komoto, H., Adachi, T., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Katoda, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flisbachman, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuchl, P., Lewis, S., Matsuo, Y., Nikaishi, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K. P., Suzuki, H., Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Borja, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlschki, S. and Hayashizaki, Y.  
RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

CONSRM  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

5  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaishi, I., Otsu, N., Saito, R., Suzuki, H., Yamada, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, B., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Gozlik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, B. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Kongaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, J. D., Reid, J., Ring, B. Z., Ringwald, M., Sander, A., Schneider, C., Sempole, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Walestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Borja, A., Yang, G. S., Yang, L., Yuan, Z., Zavanian, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komoto, H., Nakamura, K., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imocani, K., Itoh, M., Kawai, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)

CONSRM  
JOURNAL  
PUBMED

REFERENCE  
AUTHORS

6  
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Fritsch, M. C., Mada, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aldrich, V., Allen, J. E., Ameseli-Imbimbaco, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Banerji, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crome, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engelton, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemmen, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashizaki, Y., Hensch, T. K., Hirokawa, N., Hill, D., Hummel, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Ueki, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kumar, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matuzawa, S., Miki, H., Mignone, F., Miyake, S., Morise, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavese, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rosal, B., Ruan, Y., Salzberg, S. L., Sander, A., Schneider, C., Schonbach, C., Sekiguchi, K., Sempole, C. A., Sene, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugita, K., Sultana, R., Takanaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegen, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamashita, H., Zdobych, B., Zhu, S., Zimmer, A., Hide, M., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, B. T., Brusic, V., Quackenbush, J., Walestedt, C., Mattick, J. S., Hume, D. A., Kaul, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Komoto, H., Nakano, K., Nimoto, N., Nishio, T., Okada, M., Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanabe, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.  
FANTOM Consortium  
The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)

CONSRM  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

7  
Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yag, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Fritsch, M., Ravasi, T., Pang, R. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Faghihi, M. A., Sander, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Walestedt, C.  
RIKEN Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)

CONSRM  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

8 (bases 1 to 1238)  
Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, P., Iida, J., Imamura, K., Imocani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Komoto, H., Murata, N., Nakamura, M., Nishio, Y., H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail: genome-resgsc.riken.jp,  
URL: http://genome.resc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

TITLE  
JOURNAL

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
.. Division of Experimental Animal Research in Riken contributed to

prepare mouse c18ases.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.jp/>  
URL: <http://genome.gsc.riken.jp/>  
Location/Qualifiers

## FEATURES

## Source

1. 1238  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:C920016A05"  
/db\_xref="taxon:10090"  
/clone="C920016A05"  
/cell\_type="thymic cells"  
/tissue\_type="thymus"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_abbrev="2 days neonate"  
66. ->1238  
/note="unlabeled protein product; enhancer of zeste homolog  
2 (Drosophila) (MGD|WGI:107940 GB|NM\_007971, evidence:  
BLASTN, 99%, match=1181)  
putative"  
/codon\_start=1  
/protein\_id="BA32205.1"  
/db\_xref="GI:74150296"  
/translation="MGQTKKSKKSPVCMRRKRYKBYMRRLQLRFRPADEVKTMSS  
NRQILERTETLNMQRRIQPVHIMTSVSLGRTRCSVTSDDLPPADQVPLKLTN  
AVASVPIVMSPLQNMVDETVLNIIPMGDEVLDQDTFIEB.IKYNDGVHGD  
RBCGFINDRIPELVNALGOVDDDDDDDDPDEREKOKULIDNDKDETCPRKF  
PADKIPRAISMSPDKGTARELKEKYELTEQOLPGALPESCTNIDGNPAKSVORRQ  
SLSPHTLFCRCRKYDGFPLPHATPTTTRKNTERTLNDKPCGPGCTQHLBEAKRF  
AAALTAERIKTPRRPGRRGRLLPNNSRPSTPTISLESKDTSDREAGTEYGE"

## CDS

## ORIGIN

## Alignment Scores:

Pred. No.: 1.88e-179 Length: 1238  
Score: 2048.00 Matches: 381  
Percent Similarity: 99.2% Conservative: 7  
Best Local Similarity: 97.4% Mismatches: 3  
Query Match: 53.8% Indels: 0  
DB: 6 Gaps: 0

US-10-773-302-2\_copy\_1\_699 (1-699) x AKI53835 (1-1238)

QY 1 MetGlylnThrGlyLysSerGlyLysSerGlyProValCysTTPArgLysValLys 20  
DB 66 ATGGCCAGACTGGGAGAAATCTGAGAGGAGCCGGTTTGGCGGAAACGTAAAA 125  
QY 21 SerGlyLysMetArgLysGlnLysLysArgPheArgGlyLysValLysSer 40  
DB 126 TCAGAGTACATGACCTGACAGCTCAAGAGTTCAAGAGCTGATGAAGTAAAGACT 185  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnGlnLys 60  
DB 186 ATGTTTATTCATCGTCAAGAAATTTTGGAAAGACTGAACCTTAAACCAAGAGTGG 245  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyLys 80  
DB 246 AAGCAGCGGAGGATACAGCTCGTGCATCATGACTTCTGAGAGCTCATTCGCGGAGACT 305  
QY 81 ArgGlyLysSerValThrSerPheLysAspPheProThrGlnValIleProLeuLysThr 100  
DB 306 AGGAGGTGTTCATGACCACTGATGATGATTTTCCAGCAAGATCATCCCGTAAAGACC 365  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTTPSerProLeuGlnLysAsp 120  
DB 366 CTGAATGAGTGCCTCGGTGCTTAAATGATCTTTGGTGCCTTTCACACAGATTTT 425  
QY 121 MetValGlnAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAsp 140  
DB 426 ATGGTGAAGAGAACTGTTTATCATTAACATTCCTTATATGCGGAGATGAGTTCTGAT 485  
QY 141 GlnAspGlyLysPheIleGlnGlnLysLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160

DB 486 CAGATGCGACTTTCATTTGAAGAACTAATAAAAAATTATGATGGAAGAAAGTCATGCTGAC 545  
QY 161 ArgGlyLysGlyPheIleLeuAspGlnIlePheValGlnLeuValAlaValLysGln 180  
DB 546 AGGAATGTCGATTTATTAATGATGAAATTTTGTGGAGTTGGTAAATGCTCTGGGTCAA 605  
QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
DB 606 TATTAAT 665  
QY 201 LysAspLeuGlnLysAspIleAspAspAspAspAspAspAspAspAspAspAspAsp 220  
DB 666 AAGATTTAAGATTAAT 725  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLysLeu 240  
DB 726 GATTAATAATTTTGAAGCAATTCCTCAATGTTTCCAGATTAAGGACCCGCAAGAACTG 785  
QY 241 LysGlnLysTyrLysGlnLeuThrGlnGlnLysLeuProGlyLysAlaLeuProGlyLys 260  
DB 786 AAAAGAAAAATATTAAGAACTCAGGAGACAGACAGCTCCAGAGTCTGCTCTGAAATGT 845  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer 280  
DB 846 ACTCCAAACATTCATGATGACCAATGCCAATCTGTTCAAGGAGAGCAAGCTTGATTC 905  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
DB 906 TTTCAATAGCTCTTCTGTCAGCAATGTTTAAATATGATGATGATGATGATGATGATG 965  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAsnLysProCys 320  
DB 966 GCAACACCAACACTTAAAG 1025  
QY 321 GlyProGlnCysTyrGlnHisLeuGlnGlnValAlaSerLysPheAlaAlaLeuThrAla 340  
DB 1026 GAGCCAGAGTTCATTCAGCATCTGAGAGGAGCTTAAGAGTTTGTGCTGCTTACTGCT 1085  
QY 341 GlnArgIleLysThrProProLysArgProGlyLysArgArgGlyLysArgLysLeuProAsn 360  
DB 1086 GAGGTATTAAGACACACCACTTAAACGCCAGGAGGAGGAGAGAGAGAGAGAGAGAG 1145  
QY 361 AsnSerSerArgProSerSerThrProThrIleAsnValLeuGlnLysSerLysAspThrAspSer 380  
DB 1146 AACAGTACAGACCCACACCCACCATCATGATGCTGAGATCAAGATTAACAGACAGT 1205  
QY 381 AspArgGlnAlaGlyThrGlnLysGlyLysGln 391  
DB 1206 GACAG 1238

RESULT 13  
AKI64192  
LOCUS  
DEFINITION  
MUS musculus adult male hippocampus cDNA, RIKEN full-length  
enriched library, clone:G630043p15 product:enhancer of zeste  
homolog 1 (Drosophila), full insert sequence.  
AKI64192  
VERSION  
AKI64192.1 GI:74211195  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
MUS musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to



- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)
- 11042159
- 3  
Shibata,K., Itoh,M., Aizawa,K., Nagao,S., Sasaki,N., Carninci,P.,  
Komoto,H., Akiyama,T., Nishi,K., Kitanaka,T., Tashiro,H., Itoh,M.,  
Sunii,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,  
Yamanoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
Genome Res. 10 (11), 1757-1771 (2000)
- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- 4  
Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,  
Aizawa,T., Hara,A., Fukunishi,Y., Komoto,H., Adachi,J., Fukuda,S.,  
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,  
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,  
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,  
Fleischmann,M., Gaasterland,T., Gissi,C., King,B., Kochwa,H.,  
Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G.,  
Quackenbush,J., Schriml,L.M., Stebbins,F., Suzuki,R., Tomita,M.,  
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,  
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,  
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Hill,D.,  
Fletcher,C., Fujita,M., Gariboldi,M., Gustinich,S., Hill,D.,  
Hotchman,M., Hume,D.A., Kamya,M., Lee,N.H., Lyons,P.,  
Machomoni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,  
Ringo,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,  
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,  
Toyo-Oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L.,  
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,J., Kohetsuki,S.  
and Hayashizaki,Y.  
RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)
- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- 5  
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaide,I., Oosato,N., Saito,R., Suzuki,H., Yamanaka,I.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,  
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,  
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,  
Chochoja,C., Corbani,L.B., Cousins,S., Dalla,B., Dragani,T.A.,  
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,  
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,  
Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,B.D., Kanai,A.,  
Kawai,J., Kawasawa,Y., Kedzierski,R.M., King,B.L., Kongaya,A.,  
Kurouchkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,  
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,  
Ravasi,T., Reed,J.C., Reed,J.C., Reid,U., Ring,B.Z., Ringwald,M.,  
Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shmada,K.,  
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,  
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yangiawwa,M., Yang,I.,  
Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Komoto,H., Nakamura,K.,  
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
Aizawa,T., Fukuda,S., Hara,A., Hashizume,W., Imocani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Boerger,J., Birney,E. and Hayashizaki,Y.  
FANTOM Consortium  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420 (6915), 563-573 (2002)
- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- 6  
Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Fritch,M.C.,  
Medea,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C., Kodzius,R.,  
Shimokawa,K., Bajic,V.B., Bremner,S.E., Batalov,S., Forrest,A.R.,  
Zavalan,M., Davis,M.J., Wilming,L.G., Aldred,V., Allen,J.E.,  
Ambesi-Implombato,A., Apweiler,R., Attvall,Y., R.N., Bailey,T.L.,  
Bansal,M., Baxter,L., Beisel,K.W., Beran,T., Bono,H., Chalk,A.M.,  
Chiu,K.P., Choudhary,V., Christoffels,A., Clutebuck,D.R.,  
Crome,M.L., Dalla,E., Dalrymple,B.P., de Bono,B., Della Gatta,G.,  
di Bernardo,D., Doolittle,R., Engstrom,P., Fagiolini,M., Paulkner,G.,  
Fletcher,C.F., Fukushima,T., Furuno,M., Fukui,S., Gariboldi,M.,  
Georgii-Hemming,P., Gingeras,T.R., Gojobori,T., Green,R.E.,  
Gustinich,S., Harbers,M., Hayashizaki,Y., Hensch,T.K., Hirokawa,N.,  
Hill,D., Hummelbeck,L., Iacono,M., Ikeo,K., Iwama,A., Ishikawa,T.,  
Jakt,M., Kanapin,A., Katoh,M., Kawasawa,Y., Kelso,J., Kitamura,H.,  
Kitano,H., Kollas,G., Krishnan,S.P., Kruger,A., Kummerfeld,S.K.,  
Kurouchkin,I.V., Lareau,L.P., Lazarevic,D., Lipovich,L., Liu,J.,  
Ljuni,S., McWilliam,S., Madan Babu,M., Madera,M., Marchionni,L.,  
Matsuda,H., Matuzawa,S., Miki,H., Mignone,F., Miyake,S.,  
Morris,K., Mortazavi,T., Muller,N., Nakano,N., Nakachi,H.,  
Ng,P., Nilsson,R., Nishiguchi,S., Nishikawa,S., Nori,F., Ohara,O.,  
Okazaki,Y., Orlando,V., Pang,K.C., Pavan,W.J., Pavoni,G.,  
Pesole,G., Petrovsky,N., Piazza,S., Reed,J., Reid,J.F., Ring,B.Z.,  
Ringwald,M., Roset,B., Ruan,Y., Salzberg,S.L., Sandelin,A.,  
Schneider,C., Schonbach,C., Sekiguchi,K., Sempile,C.A., Seno,S.,  
Sessa,L., Sheng,Y., Shibata,Y., Shimada,H., Shimada,K., Silva,D.,  
Sinclet,B., Sperling,S., Stupka,R., Sugita,K., Sultana,R.,  
Takenaka,Y., Taki,K., Tammoja,K., Tan,S.L., Tang,S., Taylor,M.S.,  
Tegner,J., Teichmann,S.A., Ueda,H.R., van Nimwegen,B., Verardo,R.,  
Wei,C.L., Yegli,K., Yamanashi,H., Zabarovsky,E., Zhu,S., Zimmer,A.,  
Hide,W., Bult,C., Grimmond,S.M., Teasdale,R.D., Liu,B.T.,  
Brusic,V., Quackenbush,J., Wahlestedt,C., Matlick,J.S., Hume,D.A.,  
Kai,C., Sasaki,D., Tomaru,Y., Fukuda,S., Kanamori-Katayama,M.,  
Suzuki,M., Aoki,J., Aizawa,T., Iida,Y., Imamura,K., Itoh,M.,  
Kato,T., Kawai,J., Kawagashira,N., Kawasumi,T., Kojima,M.,  
Kondo,S., Komoto,H., Nakano,K., Nimochi,N., Nishio,T., Okada,M.,  
Plessey,C., Shibata,K., Shiraki,T., Suzuki,S., Tagami,M., Waki,K.,  
Watabiki,A., Okamura-Oho,Y., Suzuki,H., Kawai,J. and Hayashizaki,Y.  
FANTOM Consortium  
The transcriptional landscape of the mammalian genome  
Science 309 (5740), 1559-1563 (2005)
- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- 7  
Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakamichi,M.,  
Nakamura,M., Nishida,H., Tapp,C.C., Suzuki,M., Kawai,J., Suzuki,H.,  
Carninci,P., Hayashizaki,Y., Wells,C., Fritch,M., Ravasi,T.,  
Pang,K.C., Hallinan,J., Matlick,J., Hume,D.A., Lipovich,L.,  
Batalov,S., Engstrom,P.G., Mizuno,Y., Fagiolini,M.A., Sandelin,A.,  
Chalk,A.M., Mortazavi,T., Liang,Z., Lenhard,B. and  
Wahlestedt,C.  
RIKEN Genome Exploration Research Group  
Antisense transcription in the mammalian transcriptome  
Science 309 (5740), 1564-1566 (2005)
- JOURNAL  
PUBMED  
REFERENCE  
AUTHORS
- 8 (bases 1 to 2860)  
Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K.,  
Hori,F., Iida,Y., Imamura,K., Imocani,K., Itoh,M., Kanagawa,S.,  
Kawai,J., Kojima,M., Komoto,H., Murata,M., Nakamura,M., Nishio,Y.,  
Nishiyori,H., Nomura,K., Ohno,M., Sakazume,N., Sano,H., Sasaki,D.,  
Shibata,K., Shiraki,T., Tagami,M., Tagami,Y., Waki,K., Watabiki,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (14-APR-2004) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Shuto-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to



Db	1594	CCTGCACAGAAAGAAAAGAAAACACAGGTTGTGGCCGCACACTGACAGAAAATTGAG	14633
Qy	508	LeuylslysaaspGlySerSerAsnH1sVal1YrAsnY1rgInProCyasAPH1sProArg	527
Db	1644	CTGAAGAAGATTAAACAATTCTACACAAAGTATATACACCAACCTGTGACACACCAAG	1703
Qy	528	GlnProCyasAPPSerSerCyseProCyasVal1lealagInAsnPhCyseGlnlyAsPheCy	547
Db	1704	CGTCGGTGTGACAGACATGCCCCCTGCATCATGACCAGAACTTTTGTGAAGAAGTTCTGC	1763
Qy	548	GlnCyseSerSerGlnCyseGlnAsnArgPheProGlyCyasArgCyAsAs-AlaGlnCyAsAs	567
Db	1764	CAGTCAGACCCACAGAC-----GTTAAAGACACCTCGTT	1796
Qy	567	nThrlyGlnCyseProCyseTyLeuAlaValArgGlnCyasAPPSerProAsnLeuCyseLeuH	587
Db	1797	TCACCATGCTTACCAAGATGGTTGGCTTG---GAGTGCCACAGACTTTTCTCTGTC	1853
Qy	587	TCysGlyAlaAlaAspH1sTPAsPSSerLyAsnValSerCyAslyAsnCyseT1leG1	607
Db	1854	CCTTAGCTGCGCTGAGACCTTTAAATATCCCTTGTGCTGTGACAGGCATCTCTACAGAGAG	1913
Qy	607	nArglySerLyseLyseH1sLeuLeu	615
Db	1914	GAAAGGTGTGATCTCCCATGTTGTG	1938
RESULT 14			
LOCUS	AK045374	3639 bp	mRNA linear HTC 02-SEP-2005
DEFINITION			Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:823010C14 product:enhancer of zeste homolog 1 (Drosophila), full insert sequence.
ACCESSION	AK045374		
VERSION	AK045374.1	GI:26337298	
KEYWORDS	HTC, CAP trapper.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE			
AUTHORS	BuxleyOta, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	10349636		
REFERENCE			
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, N., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subtraktion of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED	11042159		
REFERENCE			
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuma, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, M., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaibiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Katahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.		
TITLE	RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	11076861		
REFERENCE			
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE			
AUTHORS	5 The PANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation		

JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

(Riken Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the PANTOM Consortium.  
Antisense Transcription in the Mammalian Transcriptome  
Science 309, 1564-1566 (2005)

The PANTOM Consortium, Riken Genome Exploration Research Group and  
Genome Science Group (Genome Network Project Core Group).  
The Transcriptional Landscape of the Mammalian Genome  
Science 309, 1559-1563 (2005)

8 (bases 1 to 3639)  
Adechi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furumori,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kigawa,I., Kasikawa,T.,  
Katoch,H., Kawat,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koys,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,K., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Saeki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.

Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 220-0045, Japan (E-mail:genome-res@gscc.riken.jp,  
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.jp/  
URL:http://antom.gsc.riken.jp/.  
Location/Qualifiers

FEATURES  
source

1..3639  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM,DB:B20110C14"  
/db\_xref="taxon:10090"  
/clone="B230110C14"  
/sex="male"  
/tissue\_type="corpora quadrigemina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
109..1848  
/note="unnamed protein product; enhancer of zeste homolog  
1 (Drosophila) (MGD|MG:1097695 GB|NM\_007970, evidence:  
BLASTN, 99%, match=1755)  
putative"  
/codon\_start=1  
/protein\_id="BAC32334.1"  
/db\_xref="GI:26337299"  
/translation="MDIASPTSGCTTYWKKRVKSEYMLRLQKLQANMGMKAALVLA  
NPKAVQKQIINBEKKLRVQVOPMKVSGHPPLKCTISIPGPSOIMLMSL  
NTVALPIVWSPLQONFWMEDETLNCIPMGGEVEDSTRELINNYDGKVG  
EEEMTPGSVTISDAVFLEVDALNQYSDSEBGANDPSGKDQDSGEDLPVTRKRN  
AIENGKSSSKQPPNDMIFSAIAMPNGSVGDMDKERARYLETMSDPALAPPQCTTH  
IDDPNAKSAYVRQSILSFHTLPRRCFCFKDYCDLFPHAPVNYKRKNKEIKIEPECG  
TDFCLLEGAKVEYLAMNPRLKSGGRRRHVVSASCNASASAMAETKEDSDRT  
GNMASSSEANSRCOTPTKORASPAAPQLCVLAPSPEVMTGAESLFRVHGTF  
NNCSIALLLGTTCQVRFQFAVESLIKLPLDMNPNAOKKKRKHRLMAHCRTIO  
LKQNNSSTGYNTYQPCDHPRPDSTCTCPITNTQNFCEKCCSPCKSTILLSSTGV  
VGLGVLRLFSFAP"

**Alignment Scores:**

Pred. No.:	2,839-141	Length:	3639
Score:	1645.50	Matches:	343
Percent Similarity:	63.8%	Conservative:	116
Best Local Similarity:	47.7%	Mismatches:	167
Query Match:	43.2%	Indels:	94
DB:	6	Gaps:	19

US-10-773-302-2\_COPY\_1\_699 (1-699) x AK045374 (1-3639)

OY		15	TTPAAGlyIaArgValIysSerGluTyrMetAspLeuArgGlnLeuLysArgPheArgArg	34
Db		151	TGGAAAAAGAAAGTAATCTGAATATATGCGGCTTCACAGCTCAACGGCTCCAGGCA	210
OY		35	AIAaerGIuValIysSerMetPheSerSerAsnArgInLysIleLeuGluArgThrGlu	54
Db		211	AATATGGGAGCAAAGGCTCTGTATGTGCAGAAATTTTGCAAAAGGTTCAAGAAAAAACCA	270
OY		55	IleLeuanrGlnGluTyrLysGlnArgArgIleGlnProValHisIleLeuThrserval	74
Db		271	ATCTCATATGAAGAAGTGAAGAAACTTCGTGCCAGCTGTTCACGCCAATGAAGCCCTG	330
OY		75	SerSerLeuArgGlyThrArgGluCysSerValThrSerAspLeu--AspPheProthr	93
Db		331	AGTGGGACCCTTTTCTGAANAAGTGtACCAATGAGAGATTTCGCCAGGTTGCACAGC	390
OY		94	GlnValIleProLeuLysThrLeuasnAlaValIaservalProIImetTyrsertyp	113
Db		391	CAGGATATGTGATGGGGTCTCTGAACACTGTGTGACGTGTTCCCATCATGATATCTCG	450
OY		114	SerProLeuGlnGlnAsnPhenMetValGluAspGluThrValLeuHisAsnIleProtyr	133
Db		451	TCCCTCTCCACAGAAATTCATGCTGGAAAGTVAGAGCGTTTGTCAATATTCACATC	510
OY		134	MetGlyAspGluValLeuLeuaspGlnAspGlyThrPheIleGluGlnLeuIleysantyr	153
Db		511	ATGGGTGACGAGGTGAAGAAAGATGAGACTTTCATCGAAGACTGATCAATACTAT	570
OY		154	AspGlyLysValHisGlyLysAspArgGlu-----CysGlyPheIleasnAspGlu	169
Db		571	GATGGCAAGTCCACGCTGAAGAAGAGATGATCCCTGATGTGTCTGATCGCATCT	630
OY		170	IlePheValGlnLeuValAsnAlaleuGlyGlnTyrAsnAspAspAspAspAsp----	187
Db		631	GGTGTCTGAGCTGGTGGATGCCCTCAACGATCTCATGAGAGAGAGAGACGGGAC	690
OY		188	-----AspAspGly-----AspAspProGluGlu-----ArgGlu	197
Db		691	AAAGCACCCCTCGATGGAAMGACAAAGACGACAGCAAAAGAGACTCGGTAAACAGAAA	750
OY		198	GluLysGlnLysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProkrArgLys	217
Db		751	CGAAGCGGCATGCTATCGAAGGCAACAA-----AAAGTTCACAG-----AAACAG	798
OY		218	PheProSerAspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAla	237
Db		799	TTTTCCAATGACATGATCTTCAGCCGCAATGGGCTCATGTTCTCTGAGAAATGCTCCCT	858
OY		238	GlnGlnLeuLysGlnLysLysTyrLysGlnLeuThrGlnGlnLeuProGluValAleupro	257
Db		859	GACGACATGAGAGAGAGGTATCGAAGCTGACAGAGATGTGCAGCCCCCAATCACTTCCC	918
OY		258	ProGlnCysThrProAsnIleAspGlyProAsnAlalyserServalGlnArgGlnInsar	277
Db		919	CCTGACATGACACCCCAACATGCATGGCCCCAACGCCMAATCATGTGAGGGGAGACGCT	978
OY		278	LeuHisSerPheHiseThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHis	297
Db		979	CTGCACTCTTCCACACCTTTTCTGCGGGCGGTGTTTAATATGACGCTCTCTCAC	1038
OY		298	ProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaleuAspAsn	317

[illegible]

QY 667 eashleuanaenaphevalValaaphalarthargylglaenlyslleargphe-- 686  
DB 2014 TATTTGGGGCATGC-CGTGTGTCCAGAGAGGTCAAGAGGAAATTAATCCCTGGTA 2072  
QY 687 -----AlaashiservValaenProaencystryalalyVal 699  
DB 2073 CTGGAGTTAGACAGTTGTACGCCACCATGTGGGTCTGGAGTCAATCTCAAGTCT 2127

RESULT 15  
CR990276 873 bp mRNA linear EST 28-JUN-2005  
LOCUS CR990276 RZPD no.9016 Homo sapiens cDNA clone RZPDp9016B0143 5',  
DEFINITION mRNA sequence.  
ACCESSION CR990276  
VERSION CR990276.1 GI:68284161  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
REFERENCE Hell,O., Ebert,L., Hennig,S., Henze,S., Radelof,U., Schneider,D.  
AUTHORS and Korn,B.  
TITLE Human T-Lymphocytes library  
JOURNAL Unpublished (2005)  
COMMENT RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Email: www.rzpd.de  
RZPD: RZPDp9016B0143.  
RZPDLIB: (Human T-Lymphocytes) RZPD LIB No.9016  
http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=9016 Contact:  
Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 100  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available from RZPD;  
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPDp9016B0143  
contact RZPD (product-support@rzpd.de) for further information.  
Primer name: qe3\_4, Primer sequence: CGGATTAACATTTCACACAG.  
Location/Qualifiers  
source  
1..873  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="RZPDp9016B0143"  
/tissue\_type="T-Lymphocytes"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_id="RZPD no.9016"  
/note="Vector: pQR80LSN\_cloned; Site\_1: SalI; Site\_2:  
NotI; Vector:  
http://www.rzpd.de/info/vectors/pQR80LSN\_cloned.pic.shtml  
; 1st strand cDNA was prepared from mRNA obtained from  
human T-Lymphocytes with a NotI - oligo(dT) primer [5'  
GACTAGTCTAGATCGAGCGGCGCCCTTTTCTTTTCTTTT 3']  
Double-stranded cDNA was ligated to SalI adaptors,  
digested with NotI and cloned into the NotI and SalI sites  
of the pQR80LSN\_cloned vector"

ALIGNMENT SCORES:  
Pred. No.: 5.07e-136 Length: 873  
Score: 1579.00 Matches: 290  
Percent Similarity: 99.7% Conservative: 0  
Best Local Similarity: 99.7% Mismatches: 1  
Query Match: 41.5% Indels: 0  
DB: 8 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x CR990276 (1-873)

QY 228 GUGUULEUULYSGULYSTRYLYSGUULEUTHRGUGINGINLEUPROGLYVALALEUPRO 257  
DB 1 GAAGAACTTAAGAAAGAAATATTAAGAACTCAAGCAAGCAAGCTCCCGGCGCATCTTCT 60  
QY 258 ProGUyCysThrProaenileaspolyProaenialeyservValaenProGLYUGLINSER 277  
DB 61 CCGAATGTACCCCAACATGATGACCAATGTCTTCTTCAAGAGAGCAAGC 120  
QY 278 LeuHISerPheHISerThleuPheCysArgArgCysPheUlySTRYASP CysPheLeuHIS 297  
DB 121 TTACAGCTCTTCAATAGCTTTTCTGTAGGGAGGTTTAAATGAATGATGCTTCTTACAT 180  
QY 298 ProPheHISalATHrProaenThrYLYSArgLYSantThrGLUTHAlaleuAaspn 317  
DB 181 CCTTTTCATGCAACACCCCAACCTTATATAGGGAAGAACACAGAAACGCTCTTACACAC 240  
QY 318 LysProCysGLYProGLInCysTYRGLNHISLeuGLUGLYAlalySGULPheAlaAla 337  
DB 241 AACCTTGTGACACAGGTGTACAGCATTTGAGGAGCAAGAGATTGCTGCT 300  
QY 338 LeuThralaGLUArgILEYSTRThrProBOLYsArgProGLYArgArgGLYArg 357  
DB 301 CTCACCGCTGAGCGGATTAAGACCCCAACAAAGTCCAGAGGCGCGAGAGAGAGCG 360  
QY 358 LeuProaenanserserserArgProserThrProThrILEaenValleuGLuserLYsAP 377  
DB 361 CTTCCCAATTAACAGTAGAGGCGCCGACCCCAACATTAATGTCTGAATCAAAAGAT 420  
QY 378 ThrAspSerAspArgGLUAlaGLYThrGLUTHrGLYGLYUaenAspLYSGULGU 397  
DB 421 ACAGACAGTAGTAGGAAAGCAAGGAGCTGAACCGGGGAGAGAAATATTAAGAAAGAA 480  
QY 398 GUGUGULYLYsAAspGLUTHrSerserserserGLUAlaAensArgCysGLINThrPro 417  
DB 481 GAAGAGAAAGAAAGTAAACTTCAGAGCTCTCTGAAGCAATTCCTGGTGTCAACACCA 540  
QY 418 IleySmetLYsProaenileGLUPROBOLYsUaenValGLUTPserGLYAlaGLUAla 437  
DB 541 ATTAAGATGAAGCAATATTAAGAACTCTGAGAAATGTGAGAGTGAGTGCTGACACC 600  
QY 438 SerMetPheArgValleuILEGLYThrTYRThraspAspPheCysAlaILEAlaArgLeu 457  
DB 601 TCAATGTTAGAGTCTCATTTGSCACTTACTATGACAAATTTCTGTGCCATTTGCTAGGTTA 660  
QY 458 IlegLYThrLYSTRCysArgGLInValTYRGLUPheArgValLYSGUserSerILElle 477  
DB 661 ATTGGAGACCAAAACATGTAGACAGGTGTATAGTTTAAAGTCAAAAGATCTAGCATCTTA 720  
QY 478 AlaProAlaProAlaGLUAspValAspThrProBOLYsLYsLYsArgLYsHISArg 497  
DB 721 GCTCAGAGCTCCCGTGAAGATGTGAATCTCTCCCAAGAGAAAGAGAGAGCAACCGG 780  
QY 498 LeuThrAlaAlaHISArgLYsArgLYsILEGLInleuLYsLYsAAspLYsSerSerPheHISAl 517  
DB 781 TTGGGGCTGACACTGCAAGAAATACAGCTGAAGAAAGAGCGCTCTTAAACCATGT 840  
QY 518 TyraenTYRGLNProCysAspHISProArgGLN 528  
DB 841 TACAATCTACACCTCTGTATCATCCAGGAG 873

Search completed: August 13, 2006, 17:20:29  
Job time : 6969.6 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 13, 2006, 13:21:31 ; Search time 306.689 Seconds  
(without alignments)  
6396.890 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699  
Perfect score: 3808  
Sequence: 1 MGQTKSEKBPVCMKRXVK.....KGNKIRPANHSVNPVCYAKV 699

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame.p2n.model -DEV=xlp  
-O=/abs/ABSWEB.spool/US10773302/runat.11082006.140230.16722/app.query.fasta\_1  
-DB=Issued Patents NA -QFMT=fasta -SUFFIX=p2n.rml -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-HOST=abs05p -USER=US10773302 @CGN\_1\_1\_407 @runat.11082006.140230.16722  
-NCPUs=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1.COMB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5.COMB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A.COMB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B.COMB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7.COMB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H.COMB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCYUS.COMB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP.COMB.seq:\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/R.COMB.seq:\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	2600	3	US-09-589-892B-1
2	3569.5	93.7	2476	3	US-10-104-047-1192
3	667	17.5	3148	3	US-09-699-266A-8
4	617.5	16.2	2136	3	US-09-177-249-1
5	617.5	16.2	2136	3	US-09-812-283-1
6	617.5	16.2	2136	3	US-09-071-838A-1
7	617	16.2	2315	3	US-09-061-769A-1
8	617	16.2	2799	3	US-10-176-884-16

9	608	16.0	2327	3	US-09-061-769A-3	Sequence 3, Appl1
10	593	15.6	2322	3	US-09-061-769A-4	Sequence 4, Appl1
11	563	14.8	489	3	US-09-589-892B-5	Sequence 5, Appl1
12	552	14.5	1406	3	US-09-699-266A-10	Sequence 10, Appl1
13	391	10.3	7015	3	US-09-177-249-6	Sequence 6, Appl1
14	391	10.3	7015	3	US-09-812-283-6	Sequence 6, Appl1
15	391	10.3	7015	5	US-09-071-838A-6	Sequence 6, Appl1
16	331	6.1	2652	3	US-09-248-796A-4828	Sequence 4828, Ap
17	228	6.0	11907	3	US-08-061-376-4	Sequence 4, Appl1
18	227	6.0	260	3	US-09-640-211A-49	Sequence 349, App
19	227	6.0	260	3	US-09-640-211A-591	Sequence 591, App
20	226.5	5.9	7890	3	US-09-949-016-2425	Sequence 2425, Ap
21	226	5.9	14255	2	US-08-320-559-1	Sequence 1, Appl1
22	226	5.9	14255	2	US-08-327-393-1	Sequence 1, Appl1
23	226	5.9	14255	2	US-08-306-691B-55	Sequence 55, Appl1
24	226	5.9	14255	3	US-08-545-860D-1	Sequence 1, Appl1
25	226	5.9	14255	7	PCR-US94-04496-1	Sequence 1, Appl1
26	222	5.8	1452	3	US-09-876-221-1	Sequence 1, Appl1
27	216.5	5.7	2732	3	US-09-896-892B-3	Sequence 3, Appl1
28	213	5.6	3217	4	US-10-094-749-270	Sequence 270, App
29	209.5	5.5	3102	5	US-10-131-831-9123	Sequence 9123, Ap
30	201.5	5.3	3378	3	US-09-949-016-5147	Sequence 5147, Ap
31	196	5.1	659	3	US-09-270-767-765	Sequence 765, App
32	196	5.1	1626	3	US-09-248-796A-1025	Sequence 1025, Ap
33	191.5	5.0	12078	3	US-09-799-451-835	Sequence 835, App
34	170	4.5	12308	3	US-09-854-133-422	Sequence 422, App
35	170	4.5	189560	3	US-09-949-016-17202	Sequence 17202, A
36	164	4.3	2053	3	US-09-270-767-247	Sequence 247, App
37	164	4.3	2053	3	US-09-270-767-15529	Sequence 15529, A
38	164	4.2	441	3	US-09-640-211A-384	Sequence 384, App
39	160	4.0	8241	3	US-09-902-540-798	Sequence 798, App
40	151.5	3.8	7912	3	US-09-949-016-3804	Sequence 3804, Ap
41	145	3.8	7940	3	US-09-632-033B-1	Sequence 1, Appl1
42	145	3.8	333	3	US-09-640-211A-403	Sequence 403, App
43	143.5	3.8	4209	3	US-09-248-796A-400	Sequence 400, App
44	143.5	3.8	8504	3	US-09-949-016-2430	Sequence 2430, Ap
45	143.5	3.8				

#### ALIGNMENTS

RESULT 1  
US-09-589-892B-1  
Sequence 1, Application US/09589892B  
Patent No. 668363  
GENERAL INFORMATION:  
APPLICANT: Jenuwein, Thomas  
APPLICANT: Labile, Gotz  
APPLICANT: O'Carroll, Donal  
APPLICANT: Risenhader, Frank  
APPLICANT: Rea, Stephen  
TITLE OF INVENTION: Chromatin-Regulator Genes  
FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/09/589,892B  
CURRENT FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945,988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(89)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (90)..(2330)



FEATURES:  
; NAME/KEY: 3'UTR  
; LOCATION: (2331)..(2600)  
US-09-589-892B-1

Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-589-892B-1 (1-2600)

1 MetGlyGlnThrGlyLysLysSerGlyLysGlyProValCysTrpArgLysArgValLys 20  
90 ATGGGCGGAGCTGGAGAGAAATCTGAGAGGGGACCGAGTTGTGGCGGAGAGCTGTAA 149  
21 SerGlyTrpMetArgLysLeuArgGlnLeuLysArgPheArgValAspGluValLysSer 40  
150 TCAGAGTACATGCGCATGACAGACGCTCAAGAGGTTCAAGACGATGAAGTAAAGAGT 209  
41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnLysTrp 60  
210 ATGTTAGTCCAAATCGTCAGAAATTTGGAAAGACGGAATCTTAAACCAAGAAATGG 269  
61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
270 AAACGCGAAGGATACAGCTCTGACATCTGAGCTTCTGTAGCTCATTTGGCGGAGCT 329  
81 ArgGlyCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
330 AGGAGGTGTGGGTACCAAGTGAATTTTCCAAACAAGTATCCCATTTAAAGACT 389  
101 LeuAsnAlaValAlaSerValProIleMetTrpSerTrpSerProLeuGlnAsnAsp 120  
390 CTGAATGAGGTGCTTCAGTACCCAAATGATTTCTGGTCTCCCTACAGCAAAATTT 449  
121 MetValGluAspGlyThrValIleuHisAsnIleProTrpMetGlyAspGluValLeuAsp 140  
450 ATGGTGAAGATGAACCTGTTTACATAACATCTCTTATATGGGAGATGAAGTTTAAAT 509  
141 GlnAspGlyThrPheIleGluGluLeuLysAsnTrpAspGlyLysValHisGlyAsp 160  
510 CAGATGGTACTTTCATTGAAGAACTAAATAAAATTAATGATGGGAAAGTACACGGGAT 569  
161 ArgGlyCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
570 AGAGATGTGGGTATTAATGATGAATTTTGTGGAGTGTGAATGCCCTTGTGCTCA 629  
181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluLysGln 200  
630 TATATGATGATGATGATGATGATGATGAGACGATCTCTGAAGAAAGAGAAAGACAG 689  
201 LysAspLeuGluAspHisIleArgAspAspLysGluSerArgProProArgLysPheProSer 220  
690 AAAAGTCTGGAGGATCACCGAGATGATAAGAAACCGGCCACCTCGGAAATTTCTTCT 749  
221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
750 GATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATGAAGGACACGCGAAGAAACTA 809  
241 LysGlyLysTrpLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGlyLys 260  
810 AAAGAAAATATTAAGAACTCACGAAACGACGCTCCAGGCGCATCTCTCTGAATGT 869  
261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
870 ACCCCCAACATGATGAGCAACAAATGCTAAATCTGTTCAGAGAGACAAAGCTTACACTCC 929  
281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisAspPheHis 300

930 TTTCTATACGCTTTTCTGTAGCCGATGTTTAAATATGACGTCTCTACATCCTTTTCA 989  
301 AlaThrProAsnThrTrpLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
990 GCACACCCCAACATCTTAAAGCGGAGAAACACAGAAACAGCTCTAGCAACAACTTGT 1049  
321 GlnProGlnCysTrpArgGlnHisIleLeuGluGlyValAlaLysGluPheAlaAlaLeuThrAla 340  
1050 GGACCACAGTGTTCACAGCATTTGGAGAGGACAAAGAGAGTTGTGCTGCTCTACCGCT 1109  
341 GlnArgIleLysThrProProLysArgProGlyGlyLysArgArgGlyArgLeuProAsn 360  
1110 GAGCGGATTAAGACCCCAACCAACCTCCAGAGAGCCGCGAGAAAGAGCGGCTTCCCAAT 1169  
361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
1170 AACAGTAGACAGGCCCAACACCCCAACATTAATGTCTGCAATCAAGATACAGACAGT 1229  
381 AspArgGluAlaGlyThrGluThrGlyGlyGluAsnAspLysGluGluGluLys 400  
1230 GATAGGAGAGCGAGACTGAAACGGGAGAGAAACATGATTAAGAAAGAAAGAGAG 1289  
401 LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
1290 AAAGATGAATCTTGAGCTCTCTGAAGCAATTCGGGTCAAAACCAATTAAGATG 1349  
421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
1350 AAGCCAAATATTGAACCTCCAGAAATGTGAGAGTGTGTCTGAAGCTCATATGTT 1409  
441 ArgValLeuIleGlyThrTrpTrpAspAspPheCysAlaIleAlaLysGluIleGlyThr 460  
1410 AGAGTCTCATTTGACACTTACTATGACATTTCTGTGCATTTGAGTTAATTTGGAGCC 1469  
461 LysThrCysArgGlnValTrpGluPheArgValLysGluSerSerIleIleAlaProAla 480  
1470 AAACATGATACAGGAGTATGAGTTTAAAGATTAACATTAATGATGCTCCAGCT 1529  
481 ProAlaGluAspValAspThrProProArgLysLysLysValGlyValHisGluLeuTrpAla 500  
1530 CCCCCTGAGATGAGTACTCTCCCAAGGAAAGAAAGAGAAACACCGGTTGTGGCT 1589  
501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValIleAsnTrp 520  
1590 GCACACTGCAGAAAGATACAGCTAATAAAGACGCGCTCTTAACATGTTTACAACAT 1649  
521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
1650 CAACCTGTGATCATCCAGCGCACGCTTGTACAGTTCTGTGCCCTTGTGTATGACAA 1709  
541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
1710 AATTGTTGAAAAGTTTGTCAATGTATGTCAAGTGTCAAAACCGCTTCCGGAGGC 1769  
561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTrpLeuAlaValArgGluCys 580  
1770 CGCTGCAGAGCAAGTGCAGAACCAAGAGAGTCCCGGTCTACCTGCTGTCCAGAGTGT 1829  
581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
1830 GACCTGACCTCTGTCTTAATGTGTGAGCGGCTGACCATTTGGACAGTAAATATGTGCC 1889  
601 CysLysAsnCysSerIleGlnArgGlySerLysLysIleLeuLeuLeuAlaProSerAsp 620  
1890 TGCAAGAACTGCAGATTCACGCGGCGCTCCAAAAGCAATCTAATGTGCGCACATCTGAC 1949  
621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
1950 GTGGAGGCTGGGGATTTTATCAAGATCTGTGCAAGAAATGAATTAATCTTCAGAA 2009  
641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTrpAspLys 660  
2010 TACTGTGAGAGATTAATTTCTCAAGATGACCTGACGAAGAGGAAAGTGTATGATTA 2069

QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
DB 2070 TACATGTCAGCTTTCCTGTTCACTTGAAACAATGATTTGTGTGATGCAACCGCAAG 2129  
QY 681 GtAunLyS11AArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
DB 2130 GGTAAACAATAATTCGTTTGCATCAATCTCGTAAATCCAACTGCTATGCAAAAGTT 2186

RESULT 2  
US-10-104-047-1192  
; Sequence 1192, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1192  
; LENGTH: 2476  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1192

Alignment Scores:  
Pred. No.: 0 Length: 2476  
Score: 3569.50 Matches: 659  
Percent Similarity: 94.3% Conservative: 0  
Best Local Similarity: 94.3% Mismatches: 1  
Query Match: 93.7% Indels: 39  
DB: 3 Gaps: 1

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-104-047-1192 (1-2476)

QY 1 MetGlyInThrGlyLysLeuSerGluLysGlyProValCysTrpArgLysArgValLys 20  
DB 93 ATGGCCAGACTGGGAAGAAATCTGAGAAAGGACCAAGTTGTGGCGGAAGCGTGTAAAA 152  
QY 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
DB 153 TCAGAGTCATGCGACTGAGACAGCTCAAGAGTTCAGACGACTGAACTAAATAGT 212  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGlnIleLeuAsnGlnGlnTrp 60  
DB 213 ATGTTAGTCCAAATCGTCGAAATTTTGGAAAGACGAAATCTTAAACCAAGAAATGG 272  
QY 61 LysGlnArgArgGlnGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 273 AAACGCGAAGGATACAGCTGTGCACTCTGACTTGTGTAGAGTCAATGGCGGAGCT 332  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 333 AGGGAG----- 338  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
DB 338 ----- 338  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 339 ---GCGAAGATGAAGACTGTTTACATAACATTCCTTATATGGGAGATGAAGTTTAAAT 395  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGluAsp 160  
DB 396 CAGAGTGTGACTTTCATTAAGAACTAAATAAAATTAATGATGGGAAAGTACACGGGAT 455  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnGln 180

DB 456 AGAGATGNGGGTTTATAATGATGAATAATTTTGTGAGATGGTGAATGCCCTTGATCA 515  
QY 161 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluLysGln 200  
DB 516 TATAATGATGATGAGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAAGACG 575  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220  
DB 576 AAAGATCTGAGAGATCACCGAGATGATTAAGAAACCGCCACCTCGAAATTTCTCTCT 635  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB 636 GATAAATATTTTGAAGCATTTTCTCATATGTTTCCAGATTAAGGCGACACGAGAAAGCTA 695  
QY 241 LysGluLysTyrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProGluCys 260  
DB 696 AAGGAAAAATATTAAGAACTCACCAAGACGCTCCAGGGGCGACTTCTCTTAAGT 755  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 756 ACCCCCAATGATGAGACCAATGCTTAATCTGTTCAGAGAGACCAAGCTTACACTCC 815  
QY 281 PheIleThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisIleProPheHis 300  
DB 816 TTTGATACGCTTTCTGTAGGGGAGTGTAAATATGACTGCTTCTCATCTTTGAT 875  
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
DB 876 GCAACACCAACACTATATAGCGGAAAGACAGAAACGCTCTGACAAACAAACCTTGT 935  
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGluValAlaLysGluPheAlaAlaLeuThrAla 340  
DB 936 GAGCACAGTGTATACACACTTTTGAAGGAGAAAGAGTTGTCTGCTCTCACCGCT 995  
QY 341 GluArgIleLysThrProProLysArgProGlyGlyArgArgArgGlyArgLeuProAsn 360  
DB 996 GAGCGGATTAAGACCCCAACAAAGCTCAGAGAGCGCGCAAGAGAGAGCGGTTCCCAT 1055  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
DB 1056 AACAGTACAGAGCCGACAGACCCCACTTAATGTGCGGAATCAAGAGATCAACAGT 1115  
QY 381 AspArgGluAlaGlyThrGluThrGlyGluGluAsnAspLysGluGluGluLys 400  
DB 1116 GATAGGAAAGAGGAGCTGAAGACGGGGAGAGAAACAATGATTAAGAAAGAAAGAAAG 1175  
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1176 AAAGATGAATCTTGAGCTCTCTGAGCAAAATCTCGGTGCAACACCAATTAAGATG 1235  
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
DB 1236 AAGCAATATATTGAACCTCTGAGAAATGAGATGAGTGGTGGAGCCCAATGTTT 1295  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1296 AAGATCTCATTTGGACTTACTATGACAAATTTCTGTGCAATGCTGATTAATGGAGCC 1355  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1356 AAACATGTATACAGGTGTATGAGTTTGAAGTCAAAAGAACTTAGATCATAGCTCCAGCT 1415  
QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
DB 1416 CCGCTGAGATGATGATACTCTCAAGAGAAAGAAAGAGAAACAACGGTGTGGGCT 1475  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1476 GCACATGCAAGAAAGATACACTGAATAAAGAGCGGCTCTCTTAACACAGTTTACACTAT 1535  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1536 CAACCTGTATCATCCAGCGGACGCTTGTGACAGTTCGTCCTTGTGTATGACACA 1595

QY 541 AsnBheCyGsluYpHeCyGlnCySerSerGluCyGlnAsnArgPheProGlyCys 560  
Db 1596 AATTGTTGAAAAGTTTGTCAATGTATGTCAGAGTGTCAAAACCGCTTCCGGAGTGC 1655  
QY 561 ArgCyGlyValArgInCyAsnThrLysGlnCyProCyGlyTyrLeuAlaValArgGluCys 580  
Db 1656 CGTGCAGAAAGCAGGTGCAACCAAGAGTCCGTCTACCTGGCTGTCCGAGAGTGT 1715  
QY 581 AspProAspLeuCySleuThrCyGslValAlaAspHisTyrPAspSerLysAsnValSer 600  
Db 1716 GACCTGACCTCTGTCTTACTTGTGAGCCCTGACCACTGGGACAGTAAATGTGTCC 1775  
QY 601 CysLysAsnCySerSerLysGlnArgGlySerLysHisSleuLeuLeuAlaProSerAsp 620  
Db 1776 TGCAAGAACTCAGATATTCAGCGGGCTCCAAAAGCATCTATTGTCTGGCCACTCTGAC 1835  
QY 621 ValAlaGlyTyrGlyLysPheLysAspProValGlnLysAsnGluPheLysSerGlu 640  
Db 1836 GTGCAGGCTGGGGATTTTATCAAAAGATCTGTGCAAAAATGAATTCATCTCAGAA 1895  
QY 641 TyrCyGslGlyLysLysSerGlnAspGluAlaAspArgGlyLysValTyrAspLys 660  
Db 1996 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAGAGGAAAGTGTATCATATA 1955  
QY 661 TyrMetCySerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
Db 1956 TACATGTGCACTTCTGTCTCAACTGAAACATGATTTTGTGTGATGCAACCGCAGAG 2015  
QY 681 GlnLysLysLysLysPheAlaAsnHisSerValAsnProAsnCyTyrAlaLysVal 699  
Db 2016 GGTAAACAATAATCTTTTGGCAATCATTCGGTAATTCAAAACCTGTAAGCAAAAGTT 2072

## RESULT 3

US-09-699-266A-8  
Sequence 8, Application US/09699266A  
Patent No. 6559354  
GENERAL INFORMATION:  
APPLICANT: Ma, Hongchang  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Odell, Joan T.  
APPLICANT: Orozco Jr., Emil M.  
APPLICANT: Rafaleki, J. Antoni  
TITLE OF INVENTION: TRANSCRIPTION AND GENE EXPRESSION REGULATORS  
FILE REFERENCE: B01164 US NA  
CURRENT FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: US/09/699,266A  
PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: PCT/US99/08385  
PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/083,212  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 8  
LENGTH: 3148  
TYPE: DNA  
ORGANISM: Zea mays  
US-09-699-266A-8

## Alignment Scores:

Pred. No.: 5.46e-54 Length: 3148  
Score: 667.00 Matches: 219  
Percent Similarity: 39.5% Conservative: 116  
Best Local Similarity: 25.8% Mismatches: 240  
Query Match: 17.5% Indels: 275  
DB: 3 Gaps: 36

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-699-266A-8 (1-3148)

QY 47 GlnLysLysLysLysPheLysThrGlnLysLeuAsnGlnGlnLysGlnArgGlyLysGln 66  
Db 308 CAAAGGTAAGTATTCAGATCAAAATTTGCTCACAAAAAGCAGATATATGCGCTATGCA 367

QY 67 ProValHisIle-----LeuThrSerValSerSerLeuArgGlyThrArg 81  
Db 368 CCCGTCATAGCTTGGATATTAATTCGCCGTGACAAAGATGTGTACTTTTCAAGACAAA 427  
QY 82 GluCySerValThrSerAspLeuAspPhe-----ProThrGlnValIle 96  
Db 428 GTCTTTCTCT-TCATCTAATGTATGTTTGTGTGGAATCTGTGCTCCAGAAATGATTT 486  
QY 97 -----ProLeuLysThrLeuAsnAlaValAlaSerValProIleMetTyrSerTyrSer 114  
Db 487 ATTAGACCAATTATA-----CTACCAAGATGCAAAAGCTTCCACTTATACCACTGATTA 543  
QY 115 ProLeuGlnGlnAsn-----PheMetValGlnAspGlnThrValLeuHisAsnIleProTyr 133  
Db 544 TTTTGTGACAGAAACCAAGGATGACAAAGACCAATCTGACTTGTGCGA----- 594  
QY 134 MetGlyAspGlnValLeuAspGlnAspGlyThrPheLysGlnLysLeuLysAsnTyr 153  
Db 594 ----- 594  
QY 154 AspGlyLysValHisGlyAspArgGlyCyGslYpHeLysAsnAspGluLysPheValGlu 173  
Db 595 ---CGAGGATTTACTATGATACCAAGTTGTGT-----GAAAGCTTATTT--- 636  
QY 174 LeuValAsnAlaLeuGlyGlnTyrAsnAspAspAspAspAspAspGlyAspAspPro 193  
Db 637 -----TGCAGTATATGTGAAGATGAAAGCCATTGAAGAT 669  
QY 194 GlnGlnLysArgGlnGlnLysGlnLysAspLeuGlnAspHis-----ArgAspAspLys 210  
Db 670 GAGGAGGAAAAAAGAAATTTAAACATTCGAAATCACTATTCGATGACAGTTCAA 729  
QY 211 GluSerArgProProArgLysPheProSerAspLysLysPheGlnAlaLysSerMet 230  
Db 730 GAATGTGAGC-----ATGCTGATCTGTACTGCAAGCGTACTGCAAC 774  
QY 231 PheProAspLysGlyThrAlaGlnGlnLysGlnLysTyrLysGlnLeuThrGlnGln 250  
Db 775 ATG-----GAGCGGGCTGCTGATGACATAAAGGCCAGTATGAATTCGATGTGAG 828  
QY 251 GlnLeuProGlyAlaLeuProPro-----Glu 259  
Db 829 AAAACTAAGATTTCTTGCAGAAAGGACGACGATATATGCAAGTGAAGATTTGTAC 888  
QY 260 CysThrProAsnLysAspGlyProAsnAlaLysSerValGlnArgGlnSerLeuHis 279  
Db 889 TGTCAACAAGATTGGATGCA-----GCATTTGAT 918  
QY 280 SerPheHisThrLeuPheCyArg-----ArgCySerPheLysTyrAspCyPhe 295  
Db 919 TCTTTTGAACATCTCTCTGTGACCAAGACAGATGTCTAGTGTGATTCGAAG 978  
QY 296 LeuHis-----ProPheHisAlaThrProAsnThrTyrLys 307  
Db 979 CTACATGGGTCTTCTCAAGATTTAGTATTTCT-----ACGAAABA 1020  
QY 308 ArgLysAsnThrGlnThrAlaLeuAspAsnLysProCyGslYpGlnGlnHis 327  
Db 1021 CAACCAAGCTTGAAGTGGCTGATGACAGTGTACCTGTGTATTCATTCATTAAC 1080  
QY 328 LeuGlnGlyAlaLysGlnPheAlaAlaLeuThrLysGlnArg-----Ile 343  
Db 1081 -----GCATCTGAACAGATGTGCTGTGTGTATTCATCAAGCTTTTGTGATGTT 1131  
QY 344 LysThrProPro-----LysArgProGlyGlyArgArgArg--- 355  
Db 1132 GAGGAGCCAACTCACTCATGACAGATGTATGAAACGACGACGATTCAAAATGAGAAAAG 1191  
QY 356 -----GlyArgLeuProAsnAsnSerArgProSerThrProThrIleLeu 371  
Db 1192 AACGGCTCCAGATGGAAGAGACTTAATCTCAACAAAGTGAAAGCTTCAACTGCAAGA 1251  
QY 372 ValLeu---GluSerLysAspThrAsp----- 379

```

Db      1253 GTTATCTCGAAGAGGATGATTTCGAGAGTACATCCATAAGCAATTAATCTCCACACAC 1311
Qy      379 -----
Db      1312 TCCCTAGTCCCTCAAAAGTTAAATTGGCCGAAAGTGAATCAGAAAGTTACCAAT 1371
Qy      379 -----
Db      1372 AGACGAATCGCTGAGAGAAATCTTATAGTGTGAAGAAAGCAAAAGGAAATGGCATCA 1431
Qy      380 SerAap-----ArgGluAlaGlyThrGluThrGly 389
Db      1432 TCTGATTTCAATTTTGTAGTGATATCTTTGGCAGGAGCAATGAGCTTAGCTGAT 1491
Qy      390 GlyGluAsnAspLysGlu-----Glu 397
Db      1492 ACACGAAATGAAATTAAGGAATTAATGTATCTTCACACAGAGTTCTCCAGACACAGA 1551
Qy      398 GluGluLysLysAspGluThr----- 404
Db      1552 AGTTCAAAAGAGAGTACAGCTCAATTGGAAACAGCTCAGCTTTGCTGAGGCTCAT 1611
Qy      405 --SerSerSerGluAlaAsnSerArg----- 413
Db      1612 AATGATTTCAACAGAGAAACCAATTAACCTCATTCACAGACAGATGTTACATGTTCA 1671
Qy      414 -----CysGlnThrProLysMetLysProAsn 423
Db      1672 AGGAAGAAGAAATGCTGATGATGAAATTTATGACAGCAGAGTGTAGTACCTGAGA----- 1725
Qy      424 IleGluProProGluAsnValGluThrSerGlyAlaGluAlaSerMetPheArgValLeu 443
Db      1726 -----TCATGGAAGGCAATTCAGCAGGAGCTTCTTGGAAGGA 1764
Qy      444 IleGlyThrTyrrAspAsnPheCysAlaIleAlaArg-----LeuIleGlyThrLys 461
Db      1765 TTAGAGATTTTGGAGAGACAGCTGTTTAATGCTCGAACCCTTGTGTTGGAATGAAAG 1824
Qy      462 ThrCysArgGlnValTyrrGluPheArgValLysGluSerSerIleAlaProAlaPro 481
Db      1825 ACGTCGAAGATGTTTCAATATATGATATATTAAGAAACAGACAGCGCTCGAGAGCT 1884
Qy      482 AlaGluAspValAsp-----ThrProProArgLysLys 492
Db      1885 CTTAGTGCTGATTCCTTCTTAAAGATATATTAAGGATGATGAGTACGAGTGCACAGA 1944
Qy      493 LysArgLysHisArg-----LeuTrpAlaAla--- 501
Db      1945 TCAGATATTTTGAAGGCGAGATTAAGTCCGTCGTTGAAGTACACCTGGAATCTGCA 2004
Qy      502 -----HisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrrAsn 519
Db      2005 GGTTCATCATTTCAAAAGATTAACGAAAGAAAGATCAGCCTTCGCA-----CAA 2055
Qy      520 TyrrGlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAla 539
Db      2056 TATATATCTTGTGGT---TGTCAAATCTACATCGGAAAGACAGTGTCCATGCTTTCAAAT 2112
Qy      540 GlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGly 559
Db      2113 GGGACATGTTGTGAGAAATATCTGTGGGTGTCCAAAATTTGCAAAATCGTTTTCGAGAG 2172
Qy      560 CysArgCys---LysAlaGlnCysAsnThrLysGlnCysProCysTyrrLeuAlaValArg 578
Db      2173 TGTCACCTGTGCAAAAGGCAAGTGTGCGACGCGCAATGTTCATGTTTTCGACGTGACAGG 2232
Qy      579 GluCysAspProAspLeuCysLeuThrCys-----GlyAlaAlaAsp----- 592
Db      2233 GAATGCGATCCGAGATGTTTGCAGAAACGTTGGGTGTGCTGTGATGATGATTCGGA 2292
Qy      593 -----HisTrpAspSerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySer 610

```

```

Db      2293 GTTCCAAACAGAGAGAGATTAATTAAGATCCGGAACATGAACCTGCTTAACAAC 2352
Qy      611 LysLysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrrGlyTyrPheIleLysAsp 630
Db      2353 CAACAAAGGCTTACTTACCTGGAAGATCAGATGCTCTGCTGGGAGCATCTCTCAAGAT 2412
Qy      631 ProValGlnLysAsnGluPheIleSerGluTyrrCysGlyGluIleIleSerGlnAspGlu 650
Db      2413 AGTTTACCAACATGAAATATCTGTGTAGTACTGTGGGAACTATCTCACACAAAGAA 2472
Qy      651 AlaAspArgArgGlyLysValTyrrAspLysTyrrMetCysSerPheLeuPheAsnLeuAsn 670
Db      2473 GCAGATTAACGCTGGAAGAAATATATGATGTGTGGAACCTATGCTCTTTTCAACCTGAAC 2532
Qy      671 AsnAspPheValValAspAlaThrArgLysGluAsnLysIleAsnPheAlaAsnHisSer 690
Db      2533 AATGATATGTTCTTTCAGACATACAGATGCTGTGACAGCTGAATTTGCCAACATGCTC 2592
Qy      691 ValAsnProAsnCysTyrrAlaLysVal 699
Db      2593 CCTGACCCGAATGCTATGCAAGGTT 2619

RESULT 4
US-09-177-249-1
; Sequence 1, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramon
; APPLICANT: Margosian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: 023070-06120US
; CURRENT APPLICATION NUMBER: US/09/177, 249
; EARLIER FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071, 838
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2136
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (43)..(2112)
; OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1) cDNA
US-09-177-249-1

Alignment Scores:
Pred. No.: 1.83e-49 Length: 2136
Score: 617.50 Matches: 191
Percent Similarity: 40.8% Conservative: 123
Best Local Similarity: 24.8% Mismatches: 252
Query Match: 16.2% Indels: 205
DB: 3 Gaps: 28

US-10-773-302-2_copy_1_699 (1-699) x US-09-177-249-1 (1-2136)
Qy      6 LysLysSerGluLysGlyProValCysTrpArgLysArgValLysSerGluTyrrMetArg 25
Db      17 AGAAGAAAGAGAGGCGAGTGTAAATGAGAGAGAAACCATGAGAGCATGCTGAGG 76
Qy      26 Leu---Arg-GlnLeuLysArgPheArgAlaAspGluValLysSerMetPheSerSe 44
Db      77 GTTTCACCCGCAACTAAATCATGATTAAGAGCAATTCGA---AAAGAGAGATTTCTGC 133
Qy      44 rAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrpLysGlnArgAr 64

```

Db 134 ATATCAAGAGAAAATT-----CGAGCTAGATACA--TTCCAAAGTGGG-----CTACTCAT 183  
Qy 64 glllelnProvalHislleThrservalSerSerleuArgly----- 79  
Db 184 GCTTCACACCAATCGTTGACTTAAACCAAGCCGCTGCAGAGATGATTAAGAGGA 243  
Qy 80 -----ThrsGluCyservalThrservalSerSerleuArgly 94  
Db 244 GACAAACAATCATTTTGTGCAGATGCAAAACCAACTGTCATTTCAGTCCCTCATCT 303  
Qy 94 nvalleProleuylThrsvalAval-----AlaserValProleme 110  
Db 304 GATTATATCTTTCAGAAATCAAGGTATGTTCTTGATGAGATCAAGATTAGTCTCTT 363  
Qy 110 cTysSerTrpSerProleuGlnGlnAsnRhemerValGluArgGluThrsValLeuHisAs 130  
Db 364 GAAGAAAGTGA-CCATTA-----TTTCTTGATGAGATGTTACCATTAATTAACAAAG 413  
Qy 130 nileProGlymetGluArgGluValleuArgGluAsnArgGluThrsPhele----- 146  
Db 414 TGTCAGAGCTCCAAATGTTGAGAGCTACCAAGATCCATTACATGGGCTTCACCAAAAG 473  
Qy 147 -GluGluLeuIleuysanGlyArgGlyValHisGlyAsnArgGlyCysGlyPhele 166  
Db 474 TAGCAGAGCTAGTGGCTGAAGATGAT--TCTGTGATGTTGAAGAGCAAAATCTATATT 530  
Qy 166 eAsnArgGluIlePheValGluLeuValAsnAlaLeuGlyGlnThrsAsnArgAsnArgAs 186  
Db 531 GAATGGTAG-----GCACTAACAATTTGACGACGTGAAGAGAGGA 569  
Qy 186 pAsnArgGluAsnArgProGluGlu--ArgGluGluGlyGlnAsnArgGluAs 205  
Db 570 GGAAGATGAAGAAGATGAAGAAATGAAGAAATGAAGAAATGTCGATTTTTCGAGGA 629  
Qy 205 pHisArgAsnArgGlyGlySerArgProArgGlyAsnRhemerVallePheG 225  
Db 630 T-----GTAGACCG 638  
Qy 225 uAlaIleSerSerMetRhemerArgly----- 235  
Db 639 ATTATATGACGGTGGGACAGACTATGAGTTGATGATGCTGCGGCGGCTCT 698  
Qy 236 -----ThrsAlaGluIleuArgGlyGlyGlnAsnArgGlyGln 249  
Db 699 CGCCAAAGTACCTGAAAGTGAATGTTGCGACATATGAAAGATGATCAATGAATCAAGCT 758  
Qy 249 uGlnGluLeuProGluAlaLeuProGluCysThrsProAsnIleArgGlyProAsnAl 269  
Db 759 TAAGAAAT-----GATGGAATGCTGG 779  
Qy 269 aLyservalGlnArgGluGlnSerLeuHisSerPheHisPhelePhe----- 285  
Db 780 TGAAGCT-----TCTGATTTGACATCCAAAGCAATTAATCTGTTCCAGATTTTGC 833  
Qy 286 -----CysArgGlyCysPheLeuArgGlyAsnArgGlyGlnAsnArgGlyGln 301  
Db 834 TGATAGAGATGATGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 891  
Qy 301 aThrsProAsnThrsGlyArgGlyAsnThrsGluThrsAlaLeu-----AsnArgGly 318  
Db 892 -GAGCCCAAGTCTAATGATCCAGAGAAATCTAGTTGTTGAGATGAGATGAGATGAGCA 950  
Qy 318 sProGluArgGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 338  
Db 951 ACCAAGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996  
Qy 338 uThrsAlaGluArgGlyGlyThrsProGluArgGlyGlyArgGlyGlyGlyGlyGlyGlyGly 358  
Db 997 -----GCTATCATGG-----AT 1010  
Qy 358 uProAsnArgGlySerArgProSerThrsProThrsIleAsnValleuGluSerGlyAsnRth 378

Db 1011 GATATATGATTAATCT----- 1026  
Qy 378 rAspSerArgGluAlaGlyThrsGluThrsGlyGluAsnAsnArgGlyGluGlu 398  
Db 1026 ----- 1026  
Qy 398 uGluGlyAsnArgGluThrsSerSerSerGluAlaAsnSerArgCysGlnThrsProI 418  
Db 1027 -----AT 1028  
Qy 418 eLysMetLysProAsnIleGluProGluAsnValGluThrsSerGlyAlaGluAlaSe 438  
Db 1029 ATCAAAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088  
Qy 438 rMetPheArgValleuIleGlyThrsGlyArgGlyGlnValGlyArgAsnRhemerCysAlaIleAlaArg 456  
Db 1089 TCTTACTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148  
Qy 457 -LeuIleGlyThrsLysThrsCysArgGlnValGlyGluPheArgValleuGluSerSer 1476  
Db 1149 ACTTCGGGGCTTAAGACGTGCTAGAGATTTACATTAATGATGCGGACAAAGATCAATG 1208  
Qy 476 e----- 11 477  
Db 1209 TACTATGCTATTAGACTTAAACAAACTACAAAGACAAATGAGTTACCAAAAAGT 1268  
Qy 477 eAlaProAlaProAlaGluAsnArgValaPheThrsProArgGlyValleuArgGlyHisArg 497  
Db 1269 ATCTGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1313  
Qy 497 gLeuThrsAlaAlaHisCysArgGlyHisIleGlnLeuLysValleuArgGlySerSerHisVal 517  
Db 1314 AAAATTAATGCTGTTATCGGCTGCTTAAAGAAAACAATGATGAGAAAGATGTTTA 1373  
Qy 517 lTrpAsnTrpGlnProCysAsnHisProArgGlnProCysAsnSerSerCysProCysVal 537  
Db 1374 TAAGCATACACACATGCTACT--TGCAAGTCAAAATGTTGACAGAAAGTCCCTGTTT 1430  
Qy 537 lIleAlaGlnAsnRhemerCysGluGlyPheCysGlnCysSerSerGluCysGlnAsnArgP 557  
Db 1431 AACTCAGAAAATGCTGCGAGAAATATGCGGGTCTCAAGAGATGCAAAATGCTCTT 1490  
Qy 557 eProGlyCysArgCysLys--AlaGlnCysAsnThrsGlnGluArgProCysGlyHisVal 576  
Db 1491 TGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1550  
Qy 576 aValArgGluCysAsnProAsnRhemerCys-----LeuThrsCysGlyAlaAlaAs 592  
Db 1551 TAATCGGAATGCAATGCAATGCTTTGTGCGAGTGGTTCCTTTCGCTGAGATGGGAC 1610  
Qy 592 pHisTrpAspSerLys--AsnValSerCysLysAsnCysSerIleGlnArgGlySerLys 611  
Db 1611 TCTTGAGGAGACACAGTGCATTAATCCAAATGCAAGATGCAATGCAATGCAATGCAATGCAAT 1670  
Qy 611 sLysHisIleLeuLeuAlaProSerArgValAlaGlyThrsGlyLysPheleIleuArgP 631  
Db 1671 TAAAAAGATTTCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1730  
Qy 631 oValGlnLysAsnGluPheIleSerGluTrpCysGlyGluIleIleSerGlnAsnArgGluAl 651  
Db 1731 TCTTAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1790  
Qy 651 aAspArgGlyGlyValGlyValGlyArgLysTrpMetCysSerPheLeuPheAsnLysAsnAs 671  
Db 1791 TAATGAGGCTGGAGAAATGAAGATCGGATGTTCTTCAATCTTTCATGATGATGATGATGATGAT 1850  
Qy 671 nAspRheValValaPheAlaThrsArgGlyGlyAsnLysIleLeuRhemAlaAsnHisSerVal 691  
Db 1851 TCACTCGAAATGCAATGCTCGCGTAAAGGAAAGCAAGTTCAAAATTTCTCAATCACTGACG 1910  
Qy 691 lAsnProAsnCysTrpAlaLysVal 699  
Db 1911 AAGACTTAATGCTTAAGCAAGTTG 1935

## RESULT 5

US-09-812-283-1

Sequence 1, Application US/09812283

Patent No. 6828477

GENERAL INFORMATION:

APPLICANT: Fischer, Robert L.

APPLICANT: Ohad, Nir

APPLICANT: Kiyosue, Tomohiro

APPLICANT: Yadegari, Ramlin

APPLICANT: Margossian, Linda

APPLICANT: Harada, John

APPLICANT: Goldberg, Robert B.

TITLE OF INVENTION: The Regents of the University of California

TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit

FILE REFERENCE: 023070-086120US

CURRENT APPLICATION NUMBER: US/09/812,283

CURRENT FILING DATE: 2001-03-19

PRIOR APPLICATION NUMBER: 09/177,249

PRIOR FILING DATE: 1998-10-22

PRIOR APPLICATION NUMBER: US 09/071,838

PRIOR FILING DATE: 1998-05-01

NUMBER OF SEQ ID NOS: 324

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 1

LENGTH: 2136

TYPE: DNA

ORGANISM: Arabidopsis sp.

FEATURES:

NAME/KEY: CDS

LOCATION: (43)..(2112)

OTHER INFORMATION: fertilization-independent endosperm 1 (FIB1) CDNA

US-09-812-283-1

## Alignment Scores:

Pred. No.: 1,836-49

Score: 617.50

Percent Similarity: 40.86

Best Local Similarity: 24.86

Query Match: 16.24

DB: 3

Length: 2136

Matches: 191

Conservative: 123

Mismatches: 252

Indels: 205

Gaps: 28

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-812-283-1 (1-2136)

QY 6 LyslySerGluYsileuGluArgThrGluLeuValLysSerGluTyrMetArg 25  
 DB 17 AGAAAAAAGAGAGCGGAGTGAATGAGAGAGAAACCATGAGCATGCTGAGG 76  
 QY 26 Leu--Arg-GlnLeuLysArgPheArgArgLysArgLysValLysSerMetPheSerSe 44  
 DB 77 GTTGGCCACCGCACTTAATCAGATMAAAGCAATGCA--AAAGAGAGATTTCTGC 133  
 QY 44 rAsnArgGlnLysIleLeuGluArgThrGluLeuValLysSerMetPheSerSe 64  
 DB 134 ATATCAAGAGAAATT-----CGAGCTGAGATACA-TTCCAAGTGTG---CTACTCAT 183  
 QY 64 gileGlnProValHisIleLeuThrSerValSerSerLeuArgGly----- 79  
 DB 184 GCTTCACACCATCATCGTTGACTTAACCAAGCCCGCTGCAGAGATGATATGAGAGA 243  
 QY 80 -----ThrArgGluCysSerValThrSerAspLeuAspPheProThrG1 94  
 DB 244 GACAAACAATCATTTTGTGAGATGCAAAACCATTCCTGCTCATGAGCCCTCATCT 303  
 QY 94 nValIleProLeuLysThrLeuAsnAlaVal-----AlaSerValProIleMe 110  
 DB 304 GATTATATTTCTTACGAAATCAAGCTTATGTTGATGAGATCAAGATTATGCTCTT 363  
 QY 110 tTyrSerTrpSerProGluGlnAsnPheMetValGluAspGluThrValLeuHisAs 130  
 DB 364 GAAGAAATGTA-CCATTA-----TTTCTTGATGAGATGTAACATTATTATCAAG 413

QY 130 nileProTyrMetGluAspGluValLeuAspGluAspGluThrPheIle----- 146  
 DB 414 TGTCAAGCTTCCATTTGTTGAGAAAGCTACAGATCCATTTACATGGGCTTCCACCAAGA 473  
 QY 147 -GluGluLeuIleLysAsnTyrAspGluLysValHisGluAspArgLysGlyPheI 166  
 DB 474 TAGCAGCTGATGCGTGAAGATGAT---TCTGTGATGTTGAAGAGCAAACTATTATTT 530  
 QY 166 eAsnAspGluIlePheValGluLeuValAsnAlaLeuGluGlnTyrAsnAspAspAs 186  
 DB 531 GAATGCTGAG-----GCACTGAATTTGACGACTGAAGAGATGA 569  
 QY 186 rAspAspAspGluAspAspProGluGlu--ArgGluGluLysGluLysAspLeuAs 205  
 DB 570 GAAAGATGAAGAGAGATGAGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 629  
 QY 205 rHisArgAspAspLysGluSerArgProProArgLysPheProSerAspLysIlePheG1 225  
 DB 630 T-----GTAGACCG 638  
 QY 225 uAlaIleSerSerMetPheProAspLysGly----- 235  
 DB 639 ATTATATGACGCGTTGGCAGAGCTATGTTGATGATCTGTCGCGCGCTCTCT 698  
 QY 236 -----ThrAlaGluLeuLysGluLysTyrTyrGluLeuThrG1 249  
 DB 699 CGCCAGTACCTCGAAGTGAATGTTGCGACATTTGAGAAATGATCAATGAATCAAGCT 758  
 QY 249 uGlnGluLeuProGluAlaLeuProProGluCysThrProAsnIleAspGluProAsnAl 269  
 DB 759 TAAGAT-----GATGAACTGCGG 779  
 QY 269 aLysSerValGlnArgGluGlnSerLeuHisSerPheHisThrLeuPhe----- 285  
 DB 780 TGAAGCT-----TCTGATTTGACATCCAGACATCACTACTGCTTCCAGATTTTGC 833  
 QY 286 -----CysArgArgCysPheLysTyrAspCysPheLeuHisProPheHisAl 301  
 DB 834 TGATAGACGTATTTGCGCTGTCGATGATATGATGATGATGATGATGATGATGATGAT 891  
 QY 301 aThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeu-----AspAsnLys 318  
 DB 892 -GAGCCGAGCTGATGATCCAGCAAGCAAAATCAATGATTTGTTGAGATGAAGATGACA 950  
 QY 318 sProCysGlyProGlnCysTyrGlnHisLeuGluGluAlaLysGluPheAlaAlaLe 338  
 DB 951 ACCATGCGATGAGCATTTGTTACTCAAGTGAAGAGATGTCAGAGAA----- 996  
 QY 338 uThrAlaGluArgLysIleLysThrProProLysArgProGluLysArgArgGluArgLe 358  
 DB 997 ---GCTGATCATGTG-----AT 1010  
 QY 358 uProAsnAspSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThr 378  
 DB 1011 GGATATGATGATGATCT----- 1026  
 QY 378 rAspSerAspArgGluAlaGlyThrGluThrGlyGluLysAsnAspLysGluGluGlu 398  
 DB 1026 ----- 1026  
 QY 398 uGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrProI 418  
 DB 1027 -----AT 1028  
 QY 418 eLysMetLysProAsnIleGluProProGluAsnValGluTyrSerGluAlaGluAlaSe 438  
 DB 1029 ATCAACAAGATTTGTGCTCAGATCCAAACAACAATATGAGACGCTGTAAGAGAGA 1088  
 QY 438 rMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArg---- 456  
 DB 1089 TCTTACTGTAAGAGATTTGATATTTGGAGAGAAACGTTGATGATGATGATGATTAACAT 1148  
 QY 457 -LeuIleGlyThrLysThrCysArgGluValTyrGluLysPheArgValLysGluSerSerI 476

Db 1149 ACTTCGGGGCGTTAAAGACGTCCCTAAGAGATTTCACATTAACATGCGCAACAAGATCAATG 1208  
 QY 476 e-----11 477  
 Db 1209 TACTATGTCATTAGACCTTAAACAATACTACAAAGACAAATGACGTTACCAAAAAAGT 1268  
 QY 477 eAlaProAlaProAlaGluAerValAerThrProBroArgLysLysValSerGlySerIleVal 497  
 Db 1269 ATCTCGAAAAAGTAGTAGTCGGTC-----CGCAAAATATGAAAGCTCCG 1313  
 QY 497 gLeuThrAlaAlaIleHisCysArgLysIleGlnIleuLysLysValAerGlySerSerAerHisIleVal 517  
 Db 1314 AAATATCTCTCGTTATCCGCGCTGCTTTAAAGAAAAACAATGACGTGAGAAAGCTAAGTTTAA 1373  
 QY 517 LTYrAsnTYrGlnProCysAerPheIleProArgGlnProCysAerSerSerCysProCysVal 537  
 Db 1374 TAGACACTACACACCATGACACTC---TGCAGATCAAAATGTGACAGACAAATCCCTTGTTT 1430  
 QY 537 LLeuAlaGlnAerPheCysGlyLysPheCysGlnCysSerSerGlnCysGlnAerAerPhe 557  
 Db 1431 AACTCACCAAAATGTCTCGAGAAATATATGGCGGTCTCAAAAGATGTCAACAATCGCTT 1490  
 QY 557 eProGlyCysArgCysLys---AlaGlnCysAerHisArgLysGlnCysProCysTYrLeuAl 576  
 Db 1491 TCGAGAGATGTATTTGTGCATTTGGCCGATGACAAATGACAAATGACAAATGTCCTTGTTTTCTGCG 1550  
 QY 576 aValArgLysCysAerProAerPheCys-----LeuThrCysGlyAlaAlaIleVal 592  
 Db 1551 TAATCGTGAATCGCATCCAGATCTTTGTTCGAGATTGTCCTGTAGCTGTGAGATGCGAC 1610  
 QY 592 PheIleTrpAerSerLys---AerValSerCysLysAerAerCysSerIleGlnAlaGlySerLys 611  
 Db 1611 TCTTGCGTACACACACAGTCGAAATCAATGACAAAGACAAATGCAATGCAATTCCTCTTCAACCA 1670  
 QY 611 sLysHisIleuLeuAlaProSerAerPheValAlaGlyTYrGlyIlePheIleLysAerPhe 631  
 Db 1671 TAAAAAGATCTCATTTGGAAAGCTGTGATGTTTCATGATGGGTGATTTTACATGGGACTC 1730  
 QY 631 oValGlnLysAerGlnPheIleSerGlnTYrCysGlyGlnIleIleSerGlnAerPheAla 651  
 Db 1731 TCTTAAAGAAATGATGATATCTCGAGAAATATATCTGAGAACTGATCACTCATGATGAAGC 1790  
 QY 651 aAerPheArgGlyLysValTYrAerPheTYrMetCysSerPheLeuPheAerLeuAerAer 671  
 Db 1791 TAATAGCGTCGGAATAATGAAAGATCGGATGGTTCCTTCACTTACCTTTACCTTGAAGA 1850  
 QY 671 nAerPheValValAerPheAlaThrArgLysGlyAerAerValIleArgPheAlaAerHisSerVal 691  
 Db 1851 TCAGCTCGAAATCGATGCTCGCGGTAAAGAAACGATTCNAATTTCTCAATCACTCAGC 1910  
 QY 691 IAsnProAerCysTYrAlaLysVal 699  
 Db 1911 AAGACTTAATGCTTACGCCAATGTTG 1935  
  
 RESULT 6  
 ; US-09-071-838A-1  
 ; Sequence 1, Application US/09071838A  
 ; Patent No. 7029917  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Robert L.  
 ; Onad, Nit  
 ; Kiyosue, Tomohiro  
 ; Yadegari, Ramin  
 ; Margossian, Linda  
 ; Harada, John  
 ; Goldberg, Robert B.  
 ; TITLE OF INVENTION: Nucleic Acids That Control Seed and  
 ; Fruit Development in Plants  
 ; NUMBER OF SEQUENCES: 324  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor

```

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,838A
FILING DATE: 01-May-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-086100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2136 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43..2112
OTHER INFORMATION: /product="Arabidopsis
fertilization-independent
endosperm 1 (FIE1)"
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-071-838A-1

Alignment Scores:
Pred. No.: 1,836-49 Length: 2136
Score: 617.50 Matches: 191
Percent Similarity: 40.8% Conservative: 123
Best Local Similarity: 24.8% Mismatches: 252
Query Match: 16.2% Indels: 205
DB: 5 Gaps: 28

US-10-773-302-2_COPY_1_699 (1-699) x US-09-071-838A-1 (1-2136)
QY 6 LytYsSerGluLysGlyProValCyTPaGlyeArGValIysSerGluTyRMeTarG 25
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   17 AGAAAAAAGAGAGAGCGAGGTGATTATGAGAGAGAAAACCATGAGCAGATGCGAGG 76
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 26 Leu---Arg-GlnLeuLysArgPheArgTrgAlaAspGluValIysSerMetPheSer 44
   ::::|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   77 GTTTCGCCACCCCAACTTAATCAGATTAAGAGCAATTCGA---AAAGAGAGATTCTGC 133
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   44 rAsnArgGluLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrpLysGlnArg 64
   ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   134 AATATCAAGAAATT-----CGAGCTAGATACA-TTCCAAAGTGTG---CTACTCAT 183
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 64 gllglnProValHisIleLeuThiSerValSerSerLeuArgGly----- 79
   ::::|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   184 GCTTCACACCAATCATCGTTGACTTAACCAAGCCGCTGACAGAGATGAATGAGGA 243
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 80 -----ThrArgLysSerValThrSerAspLeuAspPheProThrG1 94
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   244 GACAACAATCATCTTTGTGAGAAAGCAAAACCACTTCGTCATCAAGTCTCATCT 303
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 94 nValIleProLysrThrLeuAsnIaVal-----AlaSerValProIleMe 110
   ::::|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   304 GATTATTAATCTTTACAGAGATCAAGTTATGTTCTTGATGAGGATCAAGATTATGCTCT 363
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 110 tTyrsertPsrProLeuGlnGlnAsnPheMetValGluAspGluThrValLeuHisAs 130
   :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
   364 GAAGAAGATGTA-CCATT-----TTCTGTATGAAGATGTCATATTATCAAG 413

```



130 nileProtyrMetGlyAspGluValLeuAspGlnAspGlyThrPheIle----- 146  
 414 TGTCAAGCTTCCAAATTTGTTAGAGAGCTACAGATCCATTACATGGGCTTCACCAAG 473  
 147 -GluGluLeuIleValAsnTyrAspGlyValValHisGlyAspAspGluCysGlyPhe11 166  
 474 TAGCAGAGCTGATGGCTGAAAGTGAT---TCTGTGATGTTGTTAGAGACAAATCTATTATT 530  
 166 eAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGlnTyrAsnAspAspAspAs 186  
 531 GAATGTGTAG---GCACCTAGAAATTGACAGACAGTGAAGAAATGA 569  
 186 pAspAspAspGlyAspAspProGluGlu---ArgGluGluValGlnValAspLeuGlnAs 205  
 570 GGAAGATGAAGAAAGATGAGAGAAATCAAGAAAGAAATGCGAATTTTCTGAAAGA 629  
 205 pHisArgAspAspGlySerArgProProArgLysPheProSerAspLysIlePheG1 225  
 630 T-----GTAGACCG 638  
 225 uAlaIleSerSerMetPheProAspLysGly----- 235  
 639 ATTATATGACGCGTTGGGACAGACATATGTTGATGATCTGTCGTCGCGGCTCT 698  
 236 -----ThrAlaGluGluLeuLysGlyValTyrTyrGluLeuThrG1 249  
 699 CGCCAGATGACTCGAAGTGGATGTTTCGACATATTTGAAAGATCAATGAATCAAGCT 758  
 249 uGlnGlnLeuProGlyAlaLeuProProGluCysThrProAsnIleAspGlyProAsnAl 269  
 759 TAAAGAT-----GATGGAATGCTCTGG 779  
 269 aLysSerValGlnArgGluGlnSerLeuHisSerPheHisThrLeuPhe----- 285  
 780 TGAGGCT-----TCGATTTGACATCCAAAGACAATTAATCTGCTTCAGAGATTTTGC 833  
 286 -----CysArgArgCysPheLysTyrTyrAspCysPheLeuHisProPheHisAl 301  
 834 TGATAGACGTCAATTCGCTCGTGTGATGATATTCATATTCATATTCATATTCATATTCAT 891  
 301 aThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeu-----AspAsnLys 318  
 892 -GAGCCCGAGTTAAATCAATCCAGGAAAGACAATCTGTTGTTGAGATGAAGATGACA 950  
 318 sProCysGlyProGlnCysTyrGlnHisLeuGluGluAlaLysGluPheAlaAlaLe 338  
 951 ACCATGACGTGAGCATTTGTTACCTCAAGGTGAGAGTGAAGAA----- 996  
 338 uThrAlaGluArgLysThrProProLysArgProGluGlyArgArgGlyArgLe 358  
 997 ---GCTATCATGTG---AT 1010  
 358 uProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspTh 378  
 1011 GGATATATATTAATCT----- 1026  
 378 rAspSerAspArgGluAlaGlyThrGluThrGlyGluAsnAsnAspLysGluGluG1 398  
 1026 ----- 1026  
 398 uGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGlnThrPro11 418  
 1027 -----AT 1028  
 418 eLysMetLysProAsnIleGluProProGluAsnValGluTyrSerGlyAlaGluAlaSe 438  
 1029 ATCAACAGATGTTGCTCATGATCAACAAACATATGTGAGCGCTGTAGAGAGGA 1088  
 438 rMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArg----- 456  
 1089 TCTTAACTTGAAGAAATTGAGATATTTGGGAGAAACAGTTGTGATGTTGATTAACAT 1148

457 -LeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer11 476  
 1149 ACTTGGGGGGCTTAAGAGCTGCTGAGATTTTCAATTACATGCCGGAACAGATCAATG 1208  
 476 e-----11 477  
 1209 TACTATGCTATTAAGACTTAACAAACTACACAAAGACAAATCAGTTTACCAAAAAAGT 1268  
 477 eAlaProAlaProAlaGluAspValAspThrProProArgLysLysValGlyValHisArg 497  
 1269 ATCTGAAAAGAAAGTAACTAGCTGCTG-----CGCAAAAATTCAGACTCGG 1313  
 497 gLeuThrAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisVal 517  
 1314 AAATATGCTGTTATCCGCTGCTTTAAAGAAAAACAATAGTAGAAGAACTAACTTTTA 1373  
 517 TTYrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysProCysVal 537  
 1374 TAAACATCAACACATCAGTCACT---TGCAAGTCAAAATGTGAGACAGCAATGCCCTTGT 1430  
 537 ILeAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPh 557  
 1431 AACTCAGCAAAATTCCTCGAAGATATTTGCGGGGTCTCAAGAGATTCACAAATCGCTT 1490  
 557 eProGlyCysArgCysLys---AlaGlnCysAsnThrLysGlnCysProCysTyrLeuAl 576  
 1491 TGAGAGATGATATTCGCAATTTGGCCAAATGACAAATGACAAATGACAAATGCTTTTGTCT 1550  
 576 aValArgGluCysAspProAspLeuCys-----LeuThrCysGlyAlaAlaAs 592  
 1551 TAATGTGAATGCATTCAGATCTTGTCCGAGTTGTCCTTACCTGAGAGATGGCAC 1610  
 592 pHisThrAspSerLys---AsnValSerCysLysAsnCysSerIleGlnArgGlySerLys 611  
 1611 TCTTGCTGACACACAGTGCAGCAATTCATGCAAGAAACATGCAATTCCTCTTCAACCA 1670  
 611 aLysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIlePheIleLysAspPr 631  
 1671 TAAAAAGATTCATTCAGGAAAGCTGATGATTCATGATGAGGAGGATTTTCAATGGAGCT 1730  
 631 oValGlnLysAsnGluPheIleSerGluTyrCysGlyGluLeuIleSerGlnAspGluAl 651  
 1731 TCTTAAAGATGATGATCTCCGAGATATATCTGAGAACATGATCAGTCAATGATGAGC 1790  
 651 aAspArgArgGlyLysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAs 671  
 1791 TAATGAGCGTGGAGAAATAGAAAGATCGGATTCCTTCTTACCTTAACTTAACTGAATCA 1850  
 671 nAspPheValValAspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerVal 691  
 1851 TCAGCTGCAAAATCGATGCTCCCGTTAAAGAAACGAGTTCAAAATTTCAATCACTCAAGC 1910  
 691 lAsnProAsnCysTyrAlaLysVal 699  
 1911 AAGACTTAATCTTACGCCCAAGTTG 1935  
 DB  
 RESULT 7  
 US-09-061-769A-1  
 ; Sequence 1, Application US/09061769A  
 ; Patent No. 6239327  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cold Spring Harbor Labs  
 ; TITLE OF INVENTION: No. 6239327el Seed Specific Polycomb Group Gene and Methods of  
 ; TITLE OF INVENTION: use for Same  
 ; FILE REFERENCE: mea genes  
 ; CURRENT FILING DATE: 1998-04-16  
 ; PRIOR APPLICATION NUMBER: 09/061,769  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2315

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEAWT
; NAME/KEY: CDS
; LOCATION: (14)..(2083)
; US-09-061-769A-1
```

## Alignment Scores:

Pred. No.:	2,33e-49	Length:	2315
Score:	617.00	Matches:	189
Percent Similarity:	40.4%	Conservative:	116
Best Local Similarity:	25.0%	Mismatches:	239
Query Match:	16.2%	Indels:	213
	3	Gaps:	28

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-061-769A-1 (1-2315)

```

QY 18 ArgVallySerGluTyrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGlu 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58 CGAAGTAATCAGAT-----AAAAGCAAAATCGAAAGAGAGAG 96

QY 38 VallySerMetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsn 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97 ATT-----TCGCAATATCAAGAGAAAT-----CGAGCTGAGATACA-TTC 137

QY 58 GlnGluTyrLysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeu 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 CAAAGTGG--CTACTCATGCTTCAACACATCAATCGTTGACTTAACCAAGCCCGCTG 194

QY 78 ArgGly-----ThrArgGlnCysSerValThrSer 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 195 CAGAGGATGATATAGAGAGAGACAAATCACTTTGTCGAGAAATGCAAAACCACTTC 254

QY 88 AspleuAspPheProThrGlnValIleProLeuLysThrLeuAsnAlaVal----- 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 255 GTCAATTCAGTCCCTCATCTGATTTAAATTTCTTACGAAAGTCAGATTATGTTCTTGATG 314

QY 105 ---AlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPheMetValGlu 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 315 AGGATCAAGATTAATGCTCTTGAAAGAGATGA-CAATTA-----TTCTTGATGAA 364

QY 124 AspGluThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAspGlnAspGly 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 GATGACCATTAATTACCAAGTGTCAAGCTTCAATGTTGAGAACTACCAAGATCCAT 424

QY 144 ThrPheIle-----GlnGlnLeuIleLysAsnTyrAspGlyLysValHisGly 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 ACATGGGCTTTCACCAAAAGTAGCCAGCTGATGGCTGAAAGTGAAT--TCTGTGATGGT 481

QY 160 AspArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGly 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 AAGAGACAAATCTATTAATTGTAATGCTGAG-----GCATTAAGAA 520

QY 180 GlnTyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGln---ArgGlnGlu 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 521 TTGACACAGTGAAGAGATGAGAGATGAAAGAAAGATGAGAGAAATCAAGAAAGAA 580

QY 199 LysGlnLysAspLeuGlnAspHisIleAspArgAspLysGlnSerArgProProArgLysPhe 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 581 AAATCGCAATTTTCTGAAGAT----- 601

QY 219 ProSerAspLysIlePheGlnAlaIleSerSerMetPheProAspLysGly----- 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 602 -----GTAGACCGATTTATATGACGCGTTGGCAGAGCATATGCTTTGATGAT 649

QY 236 -----ThrAlaGlnGlnLeuLysGln 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 CTGATCGTGGCGGCGTCTCGCCAAAGTACTCGAAGTGGATGTTTCGACATATATGGA 709

QY 243 LysTyrLysGlnLeuThrGlnGlnGlnLeuProGlyAlaLeuProProGlnCysThrPro 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 AGATCAATGAAGTCAAGCTTAAGAAAT----- 736
```

```

QY 263 AsnIleAspGlyProAsnAlaLysSerValGlnArgGlnSerLeuHisSerPheHis 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 737 -----GATGAAGTCTGCTGAGGCT-----TCTGATTTGACATCCAGACAAATTA 784

QY 283 ThrLeuPhe-----CysArgArgCysPheLysTyrAspCys 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 ACTGCTTTCAGAGATTTTGGCTGATAGCTGATTCGCTGCTGCTGATGATATTCGATTT 844

QY 295 PheLeuHisProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAla 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 CATATGATGATGAAGAT-----GAGCCGAGCTTATGATCCAGCAAGCAAAATTAAGTTTG 901

QY 315 Leu-----AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGlnGlyAla 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 902 TTGAGAGATGAAGATAGACAAACATGACGATGATGTTTACTCAAGATGAGAGATGTG 961

QY 332 LysGlnPheAlaAlaAlaLeuThrAlaGlnArgLysIleLysThrProProLysArgProGly 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 962 ACAGAA-----GCTGATCATGTG----- 979

QY 352 GlyArgArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsn 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 980 -----ATGGAATATGATTAATCT----- 997

QY 372 ValLeuGlnSerLysAspThrAspSerAspArgGlnAlaGlyThrGlnThrGlyGlu 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 997 ----- 997

QY 392 AsnAsnAspLysGlnGlnGlnGlnLysLysAspGluThrSerSerSerSerGluAlaAsn 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 997 ----- 997

QY 412 SerArgCysGlnThrProIleLysMetLysProAsnIleGluProGlnAsnValGlu 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 998 -----ATATCAACAGAGTTGTGGCTGATCCAGTCAACACACATATG 1039

QY 432 TrpSerGlyAlaGlnAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPhe 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 TGAAGCCTGTGAAGAGATCTTACTTGAAAGAAATGAGATATTTGGAGAAACAGT 1099

QY 452 CysAlaIleAlaArg-----LeuIleGlyThrLysThrCysArgGlnValTyrGlnPhe 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1100 TGTATGTGCAATTAACATATCTTCGGGGCTTAAAGATGCTCTTAAGAGATTACATTAAC 1159

QY 470 ArgValLysGlnSerSerIle----- 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1160 ATGGCCGCAACAAAGTACATGATCATATGATTAACCTTAACAAATACACAAAGACAC 1219

QY 477 -----IleAlaProAlaProAlaGluAspValAspThrProProArg 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1220 AATCAGGTTACCAAAAAGTATCTGAAAAGTATGATGGTGGTGC----- 1264

QY 491 LysLysLysLysArgLysHisArgLeuThrAlaHisCysArgLysIleGlnLeuLysLys 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1265 CGCAAAAATGAGACCTCCGAAATATGCTGTTATCCGCTGCTTTAAAGAAAACAAT 1324

QY 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisIleProArgGlnProCys 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1325 AGTGAAGAGCAATAGTTTATTAAGACATACACATCACT-----TGCAGTCAAAATGT 1381

QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnLysPheCysGlnCysSer 550
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1382 GAGACAGCAATCCCTTGTGTTAATCAACCAAAATGCTCGCGAATAATATGCGGGTCTCA 1441

QY 551 SerGlnCysGlnAsnArgPheProGlyCysArgCysLys---AlaGlnCysAsnThrLys 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1442 AAGGATTCACAAATCGCTTGAAGATGATATGTGCAATTTGGCCAAATGACAAATGCA 1501

QY 570 GlnCysProCysTyrLeuAlaValArgGlnCysAspProAspLeuCys----- 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1502 CAATGCTCTTGTGCTGCTAATCGTAATGCGATCCAGATCTTTGTGCGAGTTGTCT 1561
```



QY 372 ValLeuGluSerLysAspThrAspSerAspArgLysLysGluThrGlnGlyGlu 391  
DB ----- 984  
QY 392 AsnAsnAspLysGluGluGluGluLysAspGluThrSerSerSerSerGluAsn 411  
DB ----- 984  
QY 412 SerArgCysGlnThrProIleLysMetLysProAsnIleGluProGluAsnValGlu 431  
DB ----- 984  
QY 985 -----ATATCAAAACAAGATTGGTCTGAGTCCAAACAACACTATG 1026  
QY 432 TrpSerGlyAlaGluAlaSerMetPheArgValLeuIleGlyThrTyrTrpAspAsn 451  
DB 1027 TGGACGCTGTAGAGAGATCTTAACTTGAAGGAATTGAGATTTGGAGAAACAGT 1086  
QY 452 CysAlaIleAlaArg-----LeuIleGlyThrLysThrCysArgGlnValTyrGluPhe 469  
DB 1087 TGTATGTTCATTAACAATATCTCGGGGCTTAAAGCTGCTTAAGATTTTACAAATTAC 1146  
QY 470 ArgValLysGluSerSerIle----- 476  
DB 1147 ATGGCGGCAACAAGATCATGTACTATGATTAAGACCTTAAACAATCTACAAAGAC 1206  
QY 477 -----IleAlaProAlaProAlaGluAspValAspThrProProArg 490  
DB 1207 AATCAGTTACCAAAAAGTATCTCGAAAAGTAGAGTCGGTC----- 1251  
QY 491 LysLysLysValGlyLysIleArgLeuThrAlaAlaHisCysArgGlyIleGlnLysLys 510  
DB 1252 CGCAAAAATCGAGCTCGAAAATATGCTGTTACCGCTCTTAAAGAAACAAT 1311  
QY 511 AspGlySerSerAsnIleValTyrAsnTyrGlnProCysAspHisProArgGlnProCys 530  
DB 1312 AGTGAAGAAGCTAAGTTTATTAAGACATACACCATGCACT--TGCAAGTCAAAAGT 1368  
QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGluLysPheCysGlnCysSer 550  
DB 1369 GGACAGCAATGCCCTTGTTTAACTCAGAAAATTCCTCGAATAATATGCGGGTCTCA 1428  
QY 551 SerGluCysGlnAsnArgPheProGlyCysArgCysLys--AlaGlnCysAsnThrLys 569  
DB 1429 AAGGATTCACAACATCGCTTGAGAGATGTAAATTGTGCAATTGGCCAAATGCAAAATGCA 1488  
QY 570 GlnCysProCysTyrLeuAlaValArgGluCysAspProAspLysCys----- 585  
DB 1489 CAATTCCTTGTGTTGCTGCTAATGCTGATGCGATCCAGATCTTGTCCGAGTTGCT 1548  
QY 586 LeuThrCysGlyAlaAlaAspHisThrAspSerLys--AsnValSerCysLysAsnCys 604  
DB 1549 CTACAGCTGTGAGATGCACTCTTGATGAGACACAGTGCAAATGCAGAAAGCAATG 1608  
QY 605 SerIleGlnArgGlySerLysLysHisLysLeuLeuAlaProSerAspValAlaGlyThr 624  
DB 1609 CAATTCCTCTTCAAAACAATTAAGATTCATTCATTCGAAAGCTCGATGTTTCATGAGTGG 1668  
QY 625 GlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGluTyrCysGlnGlu 644  
DB 1669 GGTGCAATTACATGGGACTCTCTTAAAAAGATGAGTATCTCCGAGAAATATCTGGAGA 1728  
QY 645 IleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLysTyrMetCysSer 664  
DB 1729 CTGATCAGCTCATGATGAGCTAATGAGCGTGGGAGAAATGAGATGCGATGCTTCTCC 1788  
QY 665 PheLeuPheAsnLysAsnAspPheValValAspAlaThrArgLysGlyAsnLysIle 684  
DB 1789 TACCTCTTTACCTTAAATGATCAGCTCGAAATGATGCTCGCGTAAAGGAAACAGATTTC 1848  
QY 685 ArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
DB 1849 AAATTTTCATCACTCAGCAAGACTTAATCTGACCAAGTTG 1893

US-09-061-769A-3  
/ Sequence 3, Application US/09061769A  
/ Patent No. 6239327  
/ GENERAL INFORMATION:  
/ APPLICANT: Cold Spring Harbor Labs  
/ TITLE OF INVENTION: No. 6239327el Seed Specific Polycomb Group Gene and Methods of  
/ TITLE OF INVENTION: use for Same  
/ FILE REFERENCE: mea genes  
/ CURRENT APPLICATION NUMBER: US/09/061,769A  
/ CURRENT FILING DATE: 1998-04-16  
/ PRIOR APPLICATION NUMBER: 09/061,769  
/ PRIOR FILING DATE: 1998-04-16  
/ NUMBER OF SEQ ID NOS: 21  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 3  
/ LENGTH: 2327  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURES:  
/ NAME/KEY: variation  
/ LOCATION: (1760)..(1761)  
US-09-061-769A-3  
  
Alignment Scores:  
Pred. No.: 1,75e-48 Length: 2327  
Score: 608.00 Matches: 190  
Percent Similarity: 40.2% Conservative: 115  
Best Local Similarity: 25.0% Mismatches: 239  
Query Match: 16.0% Indels: 217  
Gaps: 29  
  
US-10-773-302-2\_copy\_1\_699 (1-699) x US-09-061-769A-3 (1-2327)  
QY 18 ArgValLysSerGluTyrMetArgLeuArgGlnLysAspArgPheArgAlaAspGlu 37  
DB 58 CGAATTAATCAGAT-----AAAAGCAAAATCGAAAAGAGAG 96  
QY 38 ValLysSerMetPheSerSerAsnArgGlnLysIleLeuGluArgThrGlnIleLeuAsn 57  
DB 97 ATT-----TTCGATATCAAGAGAAAT-----CGAGCTGAGATACA--TTC 137  
QY 58 GlnGluThrLysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeu 77  
DB 138 CAATGTGG-----CTACTCATCTTCACACCATCAATGCTTTGACTTAAACAGCCCGCTG 194  
QY 78 ArgGly-----ThrArgLysCysSerValThrSer 87  
DB 195 CAGAGATGATTAATGAGAGACAAACAATCATCTTTGTGCGAATTCGAAACCAACTTC 254  
QY 88 AspLeuAspPheProThrGlnValIleProLeuLysThrLeuAsnAlaVal----- 104  
DB 255 GTCAATTCAGGCTCATCTGATTAATTAATTTTGAAGATCAAGGTTATGTTCTTGATG 314  
QY 105 ---AlaSerValProIleMetLysSerTrpSerProLeuGlnGlnAsnPheMetValGlu 123  
DB 315 AGGATCAAGATTAATGCTTGAAGAAGATGA--CCATTA-----TTTCTTGATGAA 364  
QY 124 AspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAspGlnAspGly 143  
DB 365 GATGACCATTAATTAACAAGTCAAGCTTCCATTTGTGAGAAAGTACCAAGTCCATT 424  
QY 144 ThrPheIle-----GluGluLeuIleLysAsnTyrAspGlyLysValHisGly 159  
DB 425 ACATGGGCTTCACCAAAAAGTAGCAGGTATGGCTGAAGATGAT--TCTGTGATTTGCT 481  
QY 160 AspArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGly 179  
DB 482 AAGGACAAAATCTATTATTTGAAATGGTAG-----GCACTAAG 520  
QY 180 GlnTyrAsnAspAspAspAspAspAspGlyAspAspProGluGlu---ArgGluGlu 198  
DB 521 TTGACGACTGAAGAAGATGAGAAAGTGAAGAAGATGAGAAAGAAATCAAGAAAGAA 580

```

QY 199 LysGlnLysAspLeuGlnAspHisArgAspArgLysGlnSerArgProProArgLysPhe 218
DB 581 AATGCGAATTTTCTGAAGAT----- 601
QY 219 ProSerAspLysIlePheGlnLysIleSerSerMetPheProAspLysGly----- 235
DB 602 -----GTAGACCGAATTTAATGACGCGTTGGCAGACGATCATGTTGGATGAT 649
QY 236 -----ThralGlnLysGln 242
DB 650 CTGTCGTGCGCGCTCTCGCCAACTGCTGAACTGAACTGTTTCCGACATATTTGGA 709
QY 243 LysTyrLysGlnLeuThrGlnGlnGlnProGlyAlaLeuProProGlnLysThrPro 262
DB 710 AGATCAATGAACTCAAGCTTAAGAT----- 736
QY 263 AsnIleAspGlyProAsnAlaLysSerValGlnArgGlnSerLeuHisSerPheHis 282
DB 737 -----GATGAACTGCTGTGAGCT-----TCTGATTTGACATCCAAAGCATTA 784
QY 283 ThrLeuPhe-----CysArgArgCysPheLysTyrAspCys 294
DB 785 ACTGCTTTCAGAGATTTTGTCTGATAGACGTCAATGCGCTGTCGATGATATTCGAT 844
QY 295 PheLeuHisProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGlnHisAla 314
DB 845 CATATGCGATGAGAGAT---GAGCCGAGCTCTAAGTCCAGCGAAGACAAATCTGATT 901
QY 315 Leu-----AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGlnGlnAla 331
DB 902 TTGAGAGATGAAAGATAGACACCAAGTCAAGTGTGCTTACCTCAAGGTGAGAGTGTG 961
QY 332 LysGlnPheAlaAlaLeuThrAlaGlnArgLysIleLysThrProProLysArgProGly 351
DB 965 AACAAGAA-----GCTGATCATGTG----- 979
QY 352 GlyArgArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsn 371
DB 980 -----ATGATTAATGATTACTCT----- 997
QY 372 ValLeuGlnSerLysAspThrAspSerAspArgGlnAlaGlyThrGlnThrGlyGln 391
DB 997 ----- 997
QY 392 AsnAsnAspLysGlnGlnGlnGlnLysLysAspGlnThrSerSerSerGlnAlaAsn 411
DB 997 ----- 997
QY 412 SerArgCysGlnThrProIleLysMetLysProAsnIleGlnProProGlnLysValGln 431
DB 998 -----ATATCAACACAGATTTGTGCTCGATGCTCAACAAACACCTAATG 1039
QY 432 TrpSerGlyAlaGlnLysSerMetPheArgValLeuIleGlyThrTyrTyrAspAsn 451
DB 1040 TGGACGCTGTAGAGAGATCTTAATTGAAAGAAATGAAATTTTGGAGAAACGCT 1099
QY 452 CysAlaIleAlaArg-----LeuIleGlyThrLysThrCysArgGlnValTyrGlnPhe 469
DB 1100 TGTGATGTTGCAATTAACATCTTGGGCGCTTAAGACGTGCTAGAGATTTTCAATTA 1159
QY 470 ArgValLysGlnSerSerIle----- 476
DB 1160 ATGCGGAAACAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1219
QY 477 -----IleAlaProAlaProAlaGlnLysValAspThrProProArg 490
DB 1220 AATCAGGTTACCAAAAAAGTATCTCGAAAAAGTAGAGTGTGCTC----- 1264
QY 491 LysLysLysArgLysHisArgLeuThrAlaAlaHisCysArgLysIleGlnLeuLysLys 510
DB 1265 CCGAAAAAATCGAGACTCGGAAAAATATGCTGCTTATCGCGCTTAAAGAAACAACT 1324

```

```

QY 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnProCys 530
DB 1325 AGTGAAGAGACTAATTTTATAGACATACACCATGCACT---TGAAGTCAAAATGT 1381
QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnLysPheCysGlnCysSer 550
DB 1382 GGACAGCAATGCCCTTGTTTAACTCGAAGAAATGCTCGAGAAATATATGGCGGTGTCA 1441
QY 551 SerGlnCysGlnAsnArgPheProGlyCysArgCysLys---AlaGlnCysAsnThrLys 569
DB 1442 AAGATTCGACAAATCGCTTGGAGAGATGATATGTCATGATTTGGCCAAATGCAAAATGCA 1501
QY 570 GlnCysProCysTyrLeuAlaValArgLysCysAspProAspLeuCys----- 585
DB 1502 CAATGCTTTGTTTGTGCTGCTTAATGTCATATCCAGATCTTTGTGCGAGTTGCTCT 1561
QY 586 LeuThrCysGlyAlaAlaAspHisIleThrAspSerLys---AsnValSerCysLysAsnCys 604
DB 1562 CTAGACTGTGAGATGGACATCTTGTGTGAGACACCAAGTCAAAATCCAAATGCAAAATG 1621
QY 605 SerIleGlnArgGlySerLysLysHisIleLeuLeuAlaProSerAspValAlaGlyTyr 624
DB 1622 CAATTCCTCTTCAAAACCAATTAAGATTTCTCATTTGGAAGATGATGATGATGATGATG 1681
QY 625 GlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGlnTyrCysGlyGln 644
DB 1682 GGTGATTTACATGGGACTCTCTTAAGAAATGATGATTTCTCGGAAATATCTGAGAA 1741
QY 645 IleIle-----SerGlnAspGlnAlaAspArgArgLysValTyrAspLys 660
DB 1742 CTGATCACTCACTATGATCTCATGATGAGCTAATGAGCGTGAAGATGAGATGAGATGCG 1801
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValValAspAlaThrArgLys 680
DB 1802 ATTGCTTCTCTTCACTCTTCACTTCACTGATGATGATGATGATGATGATGATGATGAT 1861
QY 681 GlyAsnLysIleAspPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699
DB 1862 GGAACGAGTTCAAAATTTCTCATGATCTGACAGACCTAATCTGACGCAAGTTTG 1918

RESULT 10
US-09-061-769A-4
/ Sequence 4, Application US/09061769A
/ Patent No. 6239327
/ GENERAL INFORMATION:
/ APPLICANT: Cold Spring Harbor Labs
/ TITLE OF INVENTION: No. 6239327el Seed Specific Polycomb Group Gene and Methods of
/ TITLE OF INVENTION: use for Same
/ FILE REFERENCE: mea genes
/ CURRENT APPLICATION NUMBER: US/09/061,769A
/ PRIOR FILING DATE: 1998-04-16
/ PRIOR APPLICATION NUMBER: 09/061,769
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 2322
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: variation
/ LOCATION: (1756)..(1763)
/ US-09-061-769A-4

Alignment Scores:
Pred. No.: 4 94e-47 Length: 2322
Score: 593.00 Matches: 189
Percent Similarity: 40.28 Conservative: 116
Best Local Similarity: 24.98 Mismatches: 239
Query Match: 15.64 Indels: 216
DB: 3 Gaps: 29

```

US-10-773-302-2\_copy\_1\_699 (1-699) x US-09-061-769A-4 (1-2322)

Qy 18 ArgValIysSerGluTyrMetArgLeuArgGlnLeuIleuIysArgPheArgAlaAspGlu 37  
58 CGAAGTAAATCGAT-----AAAAGAGCAATCGAAAAGAGAG 96

Qy 38 ValIysSerMetPheSerSerAsnArgGlnIleuGlnIleuGlnIleuGlnIleuGln 57  
97 ATT-----TCTGCATATCAAGAGAAAT-----CGAGCTGAAGATCA--TTC 137

Qy 58 GlnGluTyrIlySerGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerIeu 77  
138 CAAGTGTGG---CTACTGCTCTTCACACCAATCAATCGTTGATTAACCAAGCCGCTG 194

Qy 78 ArgGly-----ThrArgGluCysSerValThrSer 87  
195 CAGAGCATGATTAATGAGAGAGACAAATCATTTTGTGAGAGATGAGAAACCAACTTC 254

Qy 88 AspLeuAspPheProThrGlnValIleProLeuIysThrLeuAsnAlaVal----- 104  
255 GTCAATTCAGAGCCATCATGATTAATCTTACGAAATCAAGGTATATGTTCTTGATG 314

Qy 105 ---AlaSerValProIleMetCysSerTrpSerProLeuGlnGlnAsnPheMetValGlu 123  
315 AGGATCAAGATTATGCTTGAAGAAAGTGA--CCATTA-----TTCCTGATGAA 364

Qy 124 AspGluThrValLeuHisAsnIleProTyrMetGluAspGluValLeuAspGlnAspGly 143  
365 GATGATCAATTAATTAACAAAGTGTCAAGCTTCAATGTTGAGAGCTACCAAGATCCATT 424

Qy 144 ThrPheIle-----GlnGluLeuIleIysAsnTyrAspGlyIlyValHisGly 159  
425 ACATGGGTCTTCACCAAAAGTAGCCAGCTGATGGCTGAAGATGAT---TCTGTGATTTGT 481

Qy 160 AspArgGluCysGlyPheIleAsnAspGluIlePheValGlnLeuValAsnAlaLeuGly 179  
482 AAGAGCAAAATCTATTATTGATGTGATG-----GCACTAAGA 520

Qy 180 GlnTyrAsnAspAspAspAspAspAspGlyAspAspProGluGlu---ArgGluGlu 198  
521 TTGAGCAGATGAAGAGATGAGAGATGAAGAAAGATGAAGAAATCAAGAAAGAA 580

Qy 199 LysGlnIlyAspLeuGlnAspHisValArgAspAspGlySerArgProThrArgIlyPhe 218  
581 AAATGCAATTTTCTGAAGAT----- 601

Qy 219 ProSerAspIlyHisPheGlnAlaIleSerSerMetPheProAspIly----- 235  
602 -----GTAGACCGAATTATATGACGGTTGGCGAGCACTATGTTTGGATGAT 649

Qy 236 -----ThrAlaGlnGluLeuIlyGln 242  
650 CTGCTCGTGGCGGCGTCTGCGCAAGTACCTCGAAGTGAATGTTTGGACATATTGGA 709

Qy 243 LysTyrTyrGlnLeuThrGlnGlnIleuProGluAlaLeuProProGluCysThrPro 262  
710 AGATCAATGAATCAAGCTTAAGAT----- 736

Qy 263 AsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSerPheHis 282  
737 -----GATGGACCTGCTGGTAGGCT-----TCTGATTTGACATCCAAAGCAATTA 784

Qy 283 ThrIleuPhe-----CysArgArgCysPheIlyValAspCys 294  
785 ACTGCTTTCAGAGATTTTGTGATAGAGTCAATGCGCGTGTGATGATATTCAATGTT 844

Qy 295 PheIleuHisProPheHisIleAlaThrProAsnThrTyrLysValGlyAsnThrGluThrAla 314  
845 CATATGCAATGAGAGATAT---GAGCCCGAGTCTAGATCCAGGAGAAACCAATCTGATTG 901

Qy 315 Leu-----AspAsnIlyProCysGlyProGlnCysTyrGlnHisIleuGlnIlyAla 331  
902 TTGAGATGAAGATTAACCAACCAATGAGCATTTGTTACTCTCAAGGTGAGAGTGTG 961

Qy 332 LysGlnPheAlaAlaAlaLeuThrAlaGluArgIleIysThrProProIlyAspArgProGly 351  
962 ACAGAA-----GCTGATCAGTGT----- 979

Qy 352 G1YArgArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsn 371  
980 -----ATGATTAATGATTAATCTT----- 997

Qy 372 ValLeuGluSerIlyAspThrAspSerAspArgIlyAlaGlyThrGlnIlyGlyGlu 391  
997 ----- 997

Qy 392 AsnAsnAspIlySerGlnGluGlnIlyLysValAspGluThrSerSerSerSerGluAlaAsn 411  
997 ----- 997

Qy 412 SerArgCysGlnThrProIleIysMetIysProAsnIleGluProProGluAsnValGlu 431  
998 -----ATATCAACAAAGATTTGTGCTCAATGTCGAAACAAACACTATG 1039

Qy 432 TrpSerGlyAlaGluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPhe 451  
1040 TGACGCGCTGAGAGAGATCTTTACTTGAAAGAAATGAGATATTGGGAGAAACAGT 1099

Qy 452 CysAlaIleAlaArg-----LeuIleGlyThrIlyThrCysArgGlnValTyrGluPhe 469  
1100 TGTGATGTTGCATTAACCATATCTTGGGGCTTGAAGCGTGCCTAGAGATTTTACATTAAC 1159

Qy 470 ArgValIysGluSerSerIle----- 476  
1160 ATGCGCAACAAAGATCAATGATCTATGATCTAGACCTTAACAAACTACAAAGACAC 1219

Qy 477 -----IleAlaProAlaProAlaGluAspValAspThrProProArg 490  
1220 AATCAGGTTACAAAAGATGATCTGAAAAGTAGTAGTGGCTGCTC----- 1264

Qy 491 LysIlyLysValArgLysHisValGlyLeuThrAlaAlaHisCysArgLysIleGlnLeuIlyLys 510  
1265 CGCAAAAATCGAAGCTCCGAAATATGCTGTTATCGGCTGCTTTAAAGAAAACAAT 1334

Qy 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnProCys 530  
1325 AGTGAAGAGCTAAGTTTATTAAGCATACACACATGCACT---TGCAGTCAAAATGT 1381

Qy 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnIlyPheCysGlnCysSer 550  
1382 GGAAGCAATGCCCTTGTTAATCAGCAAAATGCTGCGAGAAATATTGCGGGTGTCA 1441

Qy 551 SerGluCysGlnAsnAlaArgPheProGlyCysArgCysLys---AlaGlnCysAsnThrIly 569  
1442 AAGATTTGCAACAAATCGTTTGAAGATGTAATGTGCAATTTGGCCAAATGCA 1501

Qy 570 GlnCysProCysTyrLeuAlaValArgGluCysAspProAspLeuCys----- 585  
1502 CAATGCTCTTTTGTGCTGCTATGCTGAAGTGCATGCAATCTTTGTGCGAGTTGCT 1561

Qy 586 LeuThrCysGlyAlaAlaAspHisIleTrpAspSerIys---AsnValSerCysLysAsnCys 604  
1562 CTTAGCTGTGAGATGGCACTGTTGGTAGACACCAAGTCAAAATCCAAATGCAAAACATG 1621

Qy 605 SerIleGlnArgGlySerIlyLysIleLeuLeuAlaProSerArgValAlaGlyTyr 624  
1622 CAATTTCTCTTCAACCAATTAAGATTTCTCAAGAAAGTCAATGTTGATGATGAG 1681

Qy 625 GlyIlePheIleLysAspProValGlnIlyAsnGlnPheIleSerGluTyrCysGlyGlu 644  
1682 GGTGATTTAATGAGACTCTTTAAAGAAATGATATCTCGGAGAAATTAACAGAGAA 1741

Qy 645 IleIle-----SerGlnAspGlnAlaAspArgArgGlyLysValTyrAspIlyTyrMet 662  
1742 CTGATCACTGATGATCATGATGAAGATTAAGACGTTGGAGAAATGAAGATCGGATTTGG 1801





QY 589 -----GlyAlaAlaAspHisTrpAspSerIysAsnVal----- 599  
DB 547 TGGGATGAGTGTGAGAGTGTTCACCTTAGCGAGCCACCAAGAAAGGTATGTTACC 606  
QY 600 SerCysIysAsnCysSerIleGlnArgGlySerIysIshLeuLeuLeuAlaProSer 619  
DB 607 CATTGCGGAAACAGAAAGCTCTCTTAACACAAACAGCAAGATTTTGGGAAAATCG 666  
QY 620 AspValAlaGlyTrpGlyIlePheIleIysAspProValGlnIysAsnGlnPheIleSer 639  
DB 667 GACCTTGAGAGATGGGGCGCTTCATTAGAACCCTGTGATAGAAATGACTATCTTGA 726  
QY 640 GlnTyrCysGlyGluIleIleSerGlnAspGluAlaAspArgGlyIysValTyrAsp 659  
DB 727 GAGTACACTGGTGAATTGATTTCTCACAAGAGACGACAAACCGCGCAAAATTTATGAC 786  
QY 660 LysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArg 679  
DB 787 CGGCAAACTCTGTTCTCTTGTGATTAAATGACGATTGTATGATGATGATTCG 846  
QY 680 LysGlyAsnIysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaIysVal 699  
DB 847 AAGGGGATTAATTGAAGTTGCAATTCATCTCTCCAGCCCACTGCTATGCAAGTGG 906  
RESULT 13  
US-09-177-249-6  
Sequence 6, Application US/09177249  
Patent No. 6229064  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Chad, Nir  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadevari, Ramini  
APPLICANT: Margosian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/177,249  
CURRENT FILING DATE: 1998-10-22  
EARLIER APPLICATION NUMBER: US 09/071,838  
EARLIER FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 7015  
TYPE: DNA  
ORGANISM: Arabidopsis sp.  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(7014)  
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)..(7015)  
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)..(7013)  
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3  
US-09-177-249-6  
Alignment Scores:  
Pred. No.: 1,09e-26 Length: 7015  
Score: 391.00 Matches: 101  
Percent Similarity: 37.5% Conservative: 38  
Best Local Similarity: 27.2% Mismatches: 69  
Query Match: 10.3% Indels: 163

DB: 3 Gaps: 10  
US-10-773-302-2\_copy\_1\_699 (1-699) x US-09-177-249-6 (1-7015)  
QY 491 LysIysIysIysArgIysHisIleArgLeuTrpAlaAlaHisCysArgIysIleGlnLeuIysIys 510  
DB 4782 CGCAAAAATCGAGACTCCGAAATAATGCTGTTATCGGCTGTTAAAGAAACAACAT 4841  
QY 511 AspGlySerSerAsnHisValTyrAsnTyrGlnProCysAspHisIleProArgGlnProCys 530  
DB 4842 AGTGAAGAACTAAAGTTTATTAAGCACTACACCAAGCACT--TGCAGTCAAAATGT 4898  
QY 531 AspSerSerCysProCysValIleAlaGlnAsnPheCysGlnIysPheCys----- 547  
DB 4899 GGAACAGAAATGCCCTTGTTAATCTACGAAAATTTGCTGCAAGAAATATTCGGGTATGTC 4958  
QY 547 ----- 547  
DB 4959 ATTCAATTTTCTTAACCGGAAAGATTCATGAGATTAAATTGACATGAGTTGTATT 5018  
QY 548 -----Gln-CysSerSerGlnCysGlnAsnArgPheProGlyCysArgCysIys--Al 564  
DB 5019 TTGTTCAGGTGCTCAAGAGATTGCAACAATCGCTTGGAGATGTATTTGTGCAATTCG 5078  
QY 564 agInCysAsnThrIysGlnCysProCysTyrIleValAlaArgGluCysAspProAspIle 584  
DB 5079 CCAATGCAACAATGTCGCAATGCTCTTTGTTGCTGCTTAATGCAATGCAATGCAATCT 5138  
QY 584 uCysLeuThrCysGlyValAlaAlaAspHisTrpAsp----- 595  
DB 5139 TTGTGCAAGTTGCTCTTGTAGTAACTTCACTTCAATATCTTTATACAAATCTA 5198  
QY 596 -----SerIysAsn----- 598  
DB 5199 TAATCAAGTAATTAACAACCAAAAGCTTATAAAAAAAGCTTATATATAGCTGTGAGA 5258  
QY 599 -----ValSerCysIysAsn-CysSer----- 605  
DB 5359 TGGCACTCTTGTGAGACACCAAGTGCAAATTCAAATGCAAGAACATGCAATTCCTCTCA 5318  
QY 606 -----IleGlnArgIysSerIysIshIsh----- 613  
DB 5319 AACCAATTAAGATTAATCAACGTCAAATTCGTAACGAAATTTAAACTAATATACGA 5378  
QY 614 -----LeuLeuLeuAlaProSerAspVal 622  
DB 5379 AAGCAATTAATCAATTCCTCCGATTTACTAGATTCATCTGGAAGCTGATGTC 5438  
QY 622 IagIyTrpGlyIlePheIle----- 628  
DB 5439 ATGATGGGGTGATTAATCAATGCGTAAGCAATCATGTAATATAGAATAGTTAATNG 5498  
QY 628 ----- 628  
DB 5499 TTATTTGCGATTCATAACACTTTTTTTTTTAAATATGTTTATACCTTAGACATTA 5558  
QY 629 -----LysAspProValGlnIysAsnGlnPheIle 639  
DB 5559 AATATATTGTGTATATAGTTTGAACCGGTACGAGCTCTTAATAAAGATAGATTCG 5618  
QY 639 ergIyTyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgIysValTyrA 659  
DB 5619 GAGATATATACAGAACTGATATCTCATGTAAGCTTAATGACGCGGGAATAGAAAG 5678  
QY 659 sPlsYrMetCysSerPheLeuPheAsnLeuAn-----AsnAspPhe 673  
DB 5679 ATCGGATTTGTTCTTCTTCACTCTTACTTGATGATCAGTAACCTTCAAAATATATT 5738  
QY 674 ValVal----- 675  
DB 5739 GAAGTAACGTTTATATCATTCGCGGTTACACATCTATTCGAATCAAGTAACATTTATT 5798  
QY 676 -----AspAlaThrArgIysGlyAsnIysIleArgPheAlaAsnHisSer 690

Db 5799 TTACAGCTCGAATGATCTCTGCCCTTAAGGAAGCACTTCAATTCTCATCTCA 5858  
Qy 691 ValAsnProAsnCysTyrAlaIysVal 699  
Db 5859 GCAGACCTTAAGCTGCTACGCCAAGGTA 5885

RESULT 14  
US-09-812-283-6  
; Sequence 6, Application US/09812283  
; Patent No. 6828477  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Ohad, Nir  
; APPLICANT: Kiyosue, Tomohiro  
; APPLICANT: Yadegari, Ramon  
; APPLICANT: Margossian, Linda  
; APPLICANT: Harada, John  
; APPLICANT: Goldberg, Robert B.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
; FILE REFERENCE: 023070-086120US  
; CURRENT APPLICATION NUMBER: US/09/812,283  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/177,249  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: US 09/071,838  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 324  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 7015  
; TYPE: DNA  
; ORGANISM: Arabidopsis sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(7014)  
; OTHER INFORMATION: fertilization-independent endosperm 1 (PIE1)  
; OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1  
; NAME/KEY: CDS  
; LOCATION: (2)..(7015)  
; OTHER INFORMATION: fertilization-independent endosperm 1 (PIE1)  
; OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2  
; NAME/KEY: CDS  
; LOCATION: (3)..(7013)  
; OTHER INFORMATION: fertilization-independent endosperm 1 (PIE1)  
; OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3  
US-09-812-283-6

Alignment Scores:  
Pred. No.: 1,09e-26 Length: 7015  
Score: 391.00 Matches: 101  
Percent Similarity: 37.5% Conservative: 38  
Best Local Similarity: 27.2% Mismatches: 69  
Query Match: 10.3% Indels: 163  
Gaps: 10

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-812-283-6 (1-7015)

Qy 491 LysIysIysValGlyValHisIleuTrrAlaHisCysValGlyValIleGlnLeuIysVal 510  
Db 4782 CGCAAAATATCGAGCTCGAAATATATGCTCGTATCCGCTTAAAGAAAACAAT 4841  
Qy 511 AspGlySerSerAsnHisValIysAsnTyrGlnProCysAspHisProArgGlnProCys 530  
Db 4842 AGTGAAGAAGCTAAGTTTATAGCACTACACCAATGACCT---TGCAAGTCAAAATGT 4898  
Qy 531 AspSerSerCysProCysValIleAlaGlnAsnProCysGlyIysPheCys----- 547  
Db 4899 GGAACAGCAATGCCCTTGTAACTACGAAATATGCTCGAGAAATATATGCGGATATGTC 4958  
Qy 547 ----- 547

Db 4959 ATTCAATTTTCTTAAGCCGGAAGATCCATGATTAATTGAACATGATTTGATTT 5018  
Qy 548 -----Gln-CysSerSerGlyCysGlnAsnArgPheProGlyCysValGlyS---Al 564  
Db 5019 TTTGTTCAAGGCTCTCAAGAGATTCGACACATCGCTTTGAGAGATTAATTGCAATTGG 5078  
Qy 564 aglncysasnthrlcysglnprocysetyrlleuvalalargglsucysasproasphe 584  
Db 5079 CCNATGCAAAATCGACATATGCTTGTGTTGCTGCTATGCTGATGATGATCGATCGATCT 5138  
Qy 584 ucysleuthrcysgylalalaaasphiletrpasp----- 595  
Db 5139 TTGTGGAGTTGTCTCTTAGTACACTTTCACCTTCAATATCTCTTATACAAATCTCA 5198  
Qy 596 -----SerIysAsn----- 598  
Db 5199 TAATCAAGTAATTCAAACCAAAAGCTTATTAATAAATTTATATATAGCTGTGAGA 5258  
Qy 599 -----ValSerCysIysAsn-CysSer----- 605  
Db 5259 TGGCACTCTTGGTGAGACACAGTGCAAATCCAAATGCAAGACATGCAATTCCTCTTCA 5318  
Qy 606 -----IleGlnArgGlySerIysIysHis----- 613  
Db 5319 AACCAATTAATAAGGTAATCAACGTCAATCCGTACCGAAATTTAAACTAATTATAGCA 5378  
Qy 614 -----LeuIleuValaProSerAspVala 622  
Db 5379 AAGACATTAACTATCATTTCCCGTATTTACTAGATTTCTCATGGAAGTCTGATGTTTC 5438  
Qy 622 laagltyrpglyllephele----- 628  
Db 5439 ATGATGAGGTCGATTTTACATGCTGATGCAATCATGTAATATAGTAAGTTTAATAG 5498  
Qy 628 ----- 628  
Db 5499 TTATTGTCATTCATTAACACTTTTTTTAAATATGTTTATCTTAGACCATTA 5558  
Qy 629 -----LysAspProValGlnIysAsnGluPheIles 639  
Db 5559 AATAATTATGATGATGATGTTTGAACCGCTGAGACTCTTAAATAAGATGATGATTCG 5618  
Qy 639 erglutyrcysgylglnlelleaserglnasprglnalaaasprargglsylvaltyra 659  
Db 5619 GAGATATATATCGAGACCTGATCACTCATGATGTAAGCTTAATGAGCGTGGAGAAATAG 5678  
Qy 659 splysIysMetCysSerPheIleuPheAsnIleuAsn-----AsnAspPhe 673  
Db 5679 ATCGATTTGTTCTTCTTACCTCTTACCTTGAATGATCAGGTAACTTCAGATATATTT 5738  
Qy 674 ValVal----- 675  
Db 5739 GAAGTAAGCTTTTATATCATTCGCGGGTTACACATCTATTCGAATCAAGTAACATTTAT 5798  
Qy 676 -----AspAlaThrArgIysGlnIysIleArgPheAlaAsnHisSer 690  
Db 5799 TTACAGCTCGAATATGATGCTCGCGGTAAAGAAAGAGTTCAATTTCTCATCTCA 5858  
Qy 691 ValAsnProAsnCysTyrAlaIysVal 699  
Db 5859 GCAGACCTTAAGCTGCTACGCCAAGGTA 5885

RESULT 15  
US-09-071-838A-6  
; Sequence 6, Application US/09071838A  
; Patent No. 7029917  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Ohad, Nir  
; APPLICANT: Kiyosue, Tomohiro  
; APPLICANT: Yadegari, Ramon  
; APPLICANT: Margossian, Linda

Harada, John  
Goldberg, Robert B.  
TITLE OF INVENTION: Nucleic Acids That Control Seed and  
Fruit Development in Plants  
NUMBER OF SEQUENCES: 324  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071.838A  
FILING DATE: 01-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-086100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7015 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7014  
OTHER INFORMATION: /product= "Arabidopsis  
fertilization-independent  
endoperm 3 (FIE3) STR/polycomb  
gene genomic sequence reading frame 1"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..7015  
OTHER INFORMATION: /product= "Arabidopsis  
fertilization-independent  
endoperm 3 (FIE3) STR/polycomb  
gene genomic sequence reading frame 2"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..7013  
OTHER INFORMATION: /product= "Arabidopsis  
fertilization-independent  
endoperm 3 (FIE3) STR/polycomb  
gene genomic sequence reading frame 3"  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-071-838A-6  
Alignment Scores:  
Pred. No.: 1,09e-26 Length: 7015  
Score: 391.00 Matches: 101  
Percent Similarity: 37.5% Conservative: 38  
Best Local Similarity: 27.2% Mismatches: 69  
Query Match: 10.3% Indels: 163  
DB: 5 Gaps: 10  
US-10-773-302-2\_COPY\_1\_699 (1-699) x US-09-071-838A-6 (1-7015)  
QY 491 IydylylyyAxylyyhyhlyylyytrpAlaAlhlyCyAxylyyIleghlyylyylyy 510  
DB 4782 CGCAAAATATCGAAGCTCCGAAATATGCTGTTATCCGCTTTAAAGAAACAACT 4841

QY 511 AspGlySerSerAsnHisValIyrybntYrGlnProCyAspHisProArgGlnProCys 530  
DB 4842 AGTGAAGAACTAAAGTTTATTAAGCACTACACCAAGCACT--TGCAGTCAAAATGT 4898  
QY 531 AspserserCysProCysValIleAlaGlnAspPheCysGlnlyySpheCys----- 547  
DB 4899 GGAACGAAATGCCCTTGTTAATCACTACAGAAATTTGCTCCAGAAATTTTCGGGTATGTC 4958  
QY 547 ----- 547  
DB 4959 ATTCAATTTTCTTACGCGGAGATCCATGATTAATTTGAACATGATTTGTATTT 5018  
QY 548 -----Gln-CysSerSerGlnCysGlnAsnAspPheProGlyCyAspGlyCys---Al 564  
DB 5019 TTGTTCAGGTGCTCAAAAGATTCGCAAAATCGTTTGAGAGATGATTTGCAATTCG 5078  
QY 564 aglnCyAsnThrlyyGlnCysProCysIyryleuAlaValArgGlnCyAspProAspLe 584  
DB 5079 CCAATGCAAAATCGCAATGTCCTTGTTCGCTTAATCGTAATCCATCCAGATCT 5138  
QY 584 uCySleuthrCySglYAlaAlaAspHisTrpAsp----- 595  
DB 5139 TTGTGGAAGTTGCTCTTAAAGTAACTTCACTCAATATCTTTATACAAATTTCTA 5198  
QY 596 -----SerlyAsn----- 598  
DB 5199 TAATCAAGTAATTCAAACCAAAAGCTTATTAACAACTTAAATATAGCTGTGAGA 5258  
QY 599 -----ValSerCyslyAsn-CysSer----- 605  
DB 5359 TGGCACTCTTGAGACACACGACGCAATTCATCGAAGAACATGCAATTCCTCTTCA 5318  
QY 606 ----ileglnArgGlySerlyylyshis----- 613  
DB 5319 AACCAATTAAGGTAATCAACGTCGAATCGTACGAAATTTAAACTAATATATACGA 5378  
QY 614 -----LeuLeuLeuAlaProSerAspVala 622  
DB 5379 AAGCAATTAATCTATCATTTCCCGTATTTTATCAATTCATTCATGAAAGTCAATGTC 5438  
QY 622 laglyTTPSgIlylePheIle----- 628  
DB 5439 ATGAAATGGGTGCAATTCATGAGGTAAAGCAATCATGTAATTAAGATAGTTAATAG 5498  
QY 628 ----- 628  
DB 5499 TTATTTGTCATTCATACACTTTTATTAATATGTTTATTAATTTAGACCAATTA 5558  
QY 629 -----lyAspProValGlnlyyAsnGlnubheiles 639  
DB 5559 AATATATTGTGTGATATGTTTGAACCGCTCGAAGCTCTTAAAGAAATGATGATCTCG 5618  
QY 639 erglyTyrCySglYglnleileSerGlnAspGlnAlaAspArgArgGlylyyValIyryA 659  
DB 5619 CAGAAATTAATCTGGAAGAACTGATCACTATGATGAAGCTTAAGACGTTGGGAATAGAG 5678  
QY 659 sPlySlyrMetCysSerPheleuPheAsnleuAn-----AspAspPhe 673  
DB 5679 ATCGAATGATGTTCTTCACTCTTACCTTGAATGATCAGTAATTCAGAAATATATTT 5738  
QY 674 ValValAl----- 675  
DB 5739 GAAGTAAGTTTATATCATTCGCGGTTTACATCTATTCGAATCAAAAGTAACATTAT 5798  
QY 676 -----AspAlaThrArgGlyyGlyyAsnlyyIleArgPheAlaAsnHisSer 690  
DB 5799 TTACAGCTCGAAATCGATGCTCGCGGTAAAGAAACGATTCGAATTTCTCATCACTCA 5858  
QY 691 ValAsnProAsnCysTyrAlaIyryAl 699  
DB 5859 GCAAGACCTTAATCTGCTACGCAAGGTA 5885

Mon Aug 14 10:47:47 2006

us-10-773-302-2\_copy\_1\_699.p2n.rn1

Page 19

Search completed: August 13, 2006, 17:31:09  
Job time : 366.689 secs

---

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 13, 2006, 16:46:46 ; Search time 2055.88 Seconds

(without alignments)  
6266.694 Million cell updates/sec

Title: US-10-773-302-2\_COPY\_1\_699

Perfect score: 3808

Sequence: 1 MGQTKSEKSPVCMRKVK.....KKNKIRPANHVNPCYAKV 699

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus.model -DEV=xlh  
-O=/abs/ABSSWB.spool/US10773302/runat\_11082006\_140239\_16873/app\_query.fasta\_1  
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpsbm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs07  
-USER=US10773302@cgn 1.1 2675 -runat\_11082006\_140239\_16873 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DRV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA Main:

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq:\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq:\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq:\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11E\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808	100.0	2600	3	US-09-954-456-1157 Sequence 1157, Ap

2	3808	100.0	2600	10	US-10-843-641A-4184	Sequence 4184, Ap
3	3808	100.0	2600	10	US-10-773-302-1	Sequence 1 Appli
4	3808	100.0	2711	9	US-10-723-860-5458	Sequence 5458, Ap
5	3802	99.8	2558	6	US-10-153-668-231	Sequence 231, App
6	3802	99.8	2576	7	US-10-172-118-1043	Sequence 1043, Ap
7	3802	99.8	2576	7	US-10-295-027-709	Sequence 709, App
8	3802	99.8	2576	8	US-10-342-887-1043	Sequence 1043, Ap
9	3802	99.8	2576	10	US-10-848-755A-129	Sequence 129, App
10	3802	99.8	2576	10	US-10-504-173-78	Sequence 78, Appl
11	3802	99.8	2576	12	US-10-960-414-196	Sequence 346, App
12	3781	99.3	2512	7	US-10-210-120-95	Sequence 95, Appl
13	3781	99.3	2512	10	US-10-909-035-95	Sequence 95, Appl
14	3569.5	93.7	2476	7	US-10-104-047-1192	Sequence 1192, Ap
15	3569.5	93.7	2476	16	US-10-153-668-333	Sequence 353, App
16	2863	75.2	1960	6	US-10-764-420-1602	Sequence 1602, Ap
17	2389.5	62.7	4185	10	US-10-504-173-77	Sequence 77, Appl
18	2384.5	62.6	2594	10	US-10-115-482-55	Sequence 55, Appl
19	2212.5	58.1	2522	7	US-10-115-482-55	Sequence 14072, A
20	1959.5	51.5	2448	13	US-11-097-143-14072	Sequence 14071, A
21	1735.5	45.6	5289	13	US-11-097-143-14071	Sequence 76875, A
22	1208.5	31.7	807	8	US-10-424-559-76875	Sequence 67, Appl
23	1138	29.9	714	9	US-10-333-872A-67	Sequence 5700, Ap
24	1056	27.7	555	3	US-09-796-692-5700	Sequence 5700, Ap
25	1056	27.7	555	6	US-10-040-862-5700	Sequence 5700, Ap
26	1056	27.7	555	7	US-10-057-475B-5700	Sequence 5700, Ap
27	1056	27.7	555	7	US-10-154-884B-5700	Sequence 5700, Ap
28	1056	27.7	555	9	US-10-764-324-5700	Sequence 5700, Ap
29	776	20.4	485	3	US-09-918-995-2112	Sequence 21967, A
30	723	19.0	3162	8	US-10-739-930-2112	Sequence 2112, Ap
31	719.5	18.9	2857	9	US-10-425-114-12944	Sequence 12944, Ap
32	701.5	18.4	3030	3	US-09-906-453-3	Sequence 3, Appli
33	701.5	18.4	3030	16	US-11-230-145-3	Sequence 231, Appl
34	697.5	18.3	3061	7	US-10-310-154-231	Sequence 1796, Ap
35	695	18.3	2954	8	US-10-424-559-11171	Sequence 5422, Ap
36	679	17.8	3173	9	US-10-739-930-1796	Sequence 5422, Ap
37	679	17.8	3193	9	US-10-425-115-96409	Sequence 5422, Ap
38	678	17.8	548	3	US-09-796-692-5422	Sequence 5422, Ap
39	678	17.8	548	6	US-10-040-862-5422	Sequence 5422, Ap
40	678	17.8	548	7	US-10-057-475B-5422	Sequence 5422, Ap
41	678	17.8	548	7	US-10-154-884B-5422	Sequence 5422, Ap
42	678	17.8	548	9	US-10-764-324-5422	Sequence 5422, Ap
43	673	17.7	3007	8	US-10-437-963-43565	Sequence 43565, A
44	672.5	17.7	3274	8	US-10-437-963-3210	Sequence 3210, Ap
45	666	17.5	2100	8	US-10-437-963-43587	Sequence 43587, A

#### ALIGNMENTS

RESULT 1  
US-09-954-456-1157  
; Sequence 1157, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Car  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954, 456  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/233, 617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234, 052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234, 923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 711  
; PRIOR FILING DATE: 2000-09-27

```
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1157
; LENGTH: 2600
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-1157

Alignment Scores:
Pred. NO.: 0 Length: 2600
Score: 3808.00 Matches: 699
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 3 Indels: 0
Gaps: 0

US-10-773-302-2_COPY_1_699 (1-699) x US-09-954-456-1157 (1-2600)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20
DB 90 ATGGGCGGACGATCGGGAGAAATCTGAGAAAGGACGAGTTTGCGGAGAACGCTATAAA 149
QY 21 SerGluTrpMetArgLeuArgGlnLeuLysArgPheLeuArgAlaAspGluValLysSer 40
DB 150 TCAGAGTACATGCGCATGACAGACAGCTCAAGAGTTCAACAGCTGATGAAGTAAAGAGT 209
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60
DB 210 ATGTTTATTCATTCATCGTCAGAAATTTTGGAAAGAACGAAATTTTAAACCAAGATCG 269
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80
DB 270 AAACAGCGAAGGATACAGCTGTCGACATCTGACTTCGTGAGCTCATTTGGCGGAGACT 329
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100
DB 330 AGGAGTGTCGCTGACAGTGAAGCTGATTTTCCAAACAAAGTATCCCATTAAGAAGT 389
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120
DB 390 CTGAATGCACTTCTTCAGTACCTCAATGATGTTCTTGCTCTCCCTTACAGCAGAAATTT 449
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140
DB 450 ATGGGGAAGATGAACCTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTAAAT 509
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisIleLysPro 160
DB 510 CAGAGTGGTACTTTCATTTGAAAGAACTTAATAAAATTTATGATGGAGAAAGTACACGGGAT 569
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnLysGln 180
DB 570 AGAGAAATGGGCTTTATTAATATGATGAATTTTGTGGAGTTGTGTAAATGCCCTTGTCTAA 629
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLysGln 200
DB 630 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220
DB 690 AAAGATCTCGAGAGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTTCTTCT 749
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGluLeu 240
DB 750 GATTAATAATTTTGAAGCCATTTCTCATATGTTTCCAGATTAAGGCGACACGACAAAGACTA 809
QY 241 LysGluLysTyrLysGluLeuThrGluGlnLeuProGlyAlaLeuProProGluCys 260
```

```
DB 810 AAGGAAAAATTAAGAACTCACCGAACACAGCTCCACGCGACTTCTCTCAATGT 869
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280
DB 870 ACCCCCAACATATAGTACCAATGCTAATCTGTTTCAGAGAGAGCAATGATCACTCC 929
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300
DB 930 TTTCATAGCGCTTTTCTGTAGCGCGATGTTTAAATATATACATGCTTCTTCACTCTTTCAT 989
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320
DB 990 GCACACCCCAACATTAATTAAGCGGAAACACGAAACAGCTTACCAACAACTTGT 1049
QY 321 GlyProGlnCysTyrGlnHisIleLeuGluGlyAlaLysGluPheAlaAlaIleuThrAla 340
DB 1050 GACACCAAGTGTTCACAGCATTTTGAGAGGACAAAGAGTTTGCTGCTGCTCACCGCT 1109
QY 341 GluArgIleLysThrProProLysArgProGlyGlyValArgArgArgGlyArgLeuProAsn 360
DB 1110 GAGCGATTAAGAACCCCAACCAACCTCCAGAGCGCGACAGAACGAGCGCTTCCAAAT 1169
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380
DB 1170 AACAGTACAGGCGCCACAGACCCCACTTAATATGTGCTGGAATCAAGATACAGACAT 1229
QY 381 AspArgGluAlaGlyThrGluThrGlyGlyValAsnAsnAspLysGluGluGluLys 400
DB 1230 GATAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1289
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420
DB 1290 AAAGATGAACCTTGAGGCTCTCTGAAAGCAATCTCGGTCTAAACCAATTAAGATG 1349
QY 421 LysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440
DB 1350 AAGCCAAATATTAACCTCTCGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1409
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaAlaIleArgLeuIleGlyThr 460
DB 1410 AGAGTCTCATTTGCACTTACTATGACAAATTTCTGTGCATTTGCTAGGTTAATGGGACC 1469
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480
DB 1470 AAAACATGTAGACAGGTGATGAGTTTAAAGTCAAAAGATCTACATCATGTGCCAGCT 1529
QY 481 ProIleGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500
DB 1530 CCGCGTGAAGTGTGAATCTCTCCAGAGAAAGAAAGAGAAACACCGGTTGTGGCT 1589
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520
DB 1590 GCACACTGCAGAAAGATACAGCTGAAGAAAGACCGGCTCTTAACATGTTTAAACATAT 1649
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB 1650 CAACCTGTGATCATCCACGCGACGCTTGTACAGTTCTGTCCTTGTGTATATGACAA 1709
QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560
DB 1710 AATTTTGTGAAAGTTTGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLysAlaValArgLysCys 580
DB 1770 CCGTGCAGAGCAGCTGACAGACCAAGCAGCGCGCTCTACCTGCTGCTGCTGCTGCTGCT 1829
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600
DB 1830 GACCTGACCTCTGCTTACTTGTGAGACCGCTGACCAATGGGACAGTAAAGATGTGCTC 1889
QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysHisIleLeuLeuValProSerAsp 620
```



Db 1890 TGCAGAACTGACGATTCAGCGGGCTCCAAAACATCTATTGTCGACCATCTGAC 1949  
Qy 621 ValAlaGlyThrGlyIlePheIleIleAspProValGlnIleAsnGlnPheIleSerGln 640  
Db 1950 GGGCGAGGCTGGGGGATTTTATCAAGATCTGTGCAAGAAATATATTCATCTCAGAA 2009  
Qy 641 TTYCGAGTGTGTTTCTGTTCAAGATCTGTAAGAGGAGGAGGAGGAGGAGGAGGAGG 660  
Db 2010 TACTGTGAGAGATTTTCTGTTCAAGATCTGTAAGAGGAGGAGGAGGAGGAGGAGG 2069  
Qy 661 TTYMetCysSerPheLeuPheLeuAsnAsnAspPheValValAspAlaThrArgLys 680  
Db 2070 TACATGTGACGCTTCTGTTCAACTGAACAATGATTTGTGTGATGCAACCCGCAAG 2129  
Qy 681 GYAAsnLysIleAlaArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
Db 2130 GGTAACAAAATTCGTTTGCMAATCATCTGGTAAATCAAAAGCTATGCAAAAGTT 2186  
RESULT 2  
US-10-843-641A-4184  
; Sequence 4184, Application US/10843641A  
; Publication No. US2005006445A1  
; GENERAL INFORMATION:  
; APPLICANT: Avalon Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; FILE REFERENCE: 689290-189  
; CURRENT APPLICATION NUMBER: US/10/843,641A  
; CURRENT FILING DATE: 2004-05-12  
; PRIOR APPLICATION NUMBER: US/09/873,367  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US/09/954,531  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,436  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,832  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/964,824  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/09/967,768  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US/09/968,007  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,347  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,708  
; PRIOR FILING DATE: 2001-10-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 8447  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4184  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-843-641A-4184  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 10 Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-843-641A-4184 (1-2600)  
Qy 1 MetGlyGlnThrGlyLysLysSerGlnLysGlyProValCysTTPArgLysArgValLys 20  
Db 90 ATGGGCGAGACTGGGAGAAATCTGAGAGGAGCAAGTTTGTGGCGGAAAGCGGTGTAATA 149  
Qy 21 SerGlyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGlnValLysSer 40

Db 150 TCAGAGTACATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 209  
Qy 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgGlnIleLeuAsnGlnLysP 60  
Db 210 ATGTTAGTTCGAATCGTCAGAAATTTTGGAAAGACGGAATCTTAAACAGAAATG 269  
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 270 AAACAGCGAAGATACAGCTGTGCAATCTGACTCTGTAGCTCATTTGCGGGACT 329  
Qy 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 330 AGGAGTGTTCGTTACGACGACGACGACGACGACGACGACGACGACGACGACGACG 389  
Qy 101 LeuAsnAlaValAlaSerValProIleMetCysTyrSerTyrSerProLeuGlnIleAsn 120  
Db 390 CTGAATGACGTTCTTCAGTACCATTAATGATTTCTTGAGTCCCTCAGACAGAAATTT 449  
Qy 121 MetValGlnAspGlnThrValLeuHisAsnIleProTyrMetGlyAspGlnValLeuAsp 140  
Db 450 ATGTGTGAAAGATGAAACTGTTTACATTAATCTCTTAATGAGAGATGAAAGTTAGAT 509  
Qy 141 GlnAspGlyThrPheIleGlnLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160  
Db 510 CAGAGTGTACTTTCATTTGAGAACTTAATTAATTAATTTGATGGAGAAAGTACACGGGAT 569  
Qy 161 ArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGln 180  
Db 570 AGAGAAATGGGGTTATTAATGATGAATTTTGTGGAGTTGGTGAATGCCCTTGATCA 629  
Qy 181 TTYAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
Db 630 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
Qy 201 LysAspLeuGlnAspHisArgAspAspLysGlnSerAspProArgLysPheProSer 220  
Db 690 AAAGATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGAAATTTCTTCT 749  
Qy 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnLys 240  
Db 750 GATTAATTTTGAAGCCATTTCTCATATGTTTCCAGATTAAGGAGCAGCAGAAAGACTA 809  
Qy 241 LysGlnLysTyrLysGlnLeuThrGlnGlnLeuProGlyAlaLeuProGlyLys 260  
Db 810 AAGGAAATATTAAGAACTACCGACAGCAGCTCCAGCGCAGCTTCTCTGAATGT 869  
Qy 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnSerLeuHisSer 280  
Db 870 ACCCCCAACATGATGACCAAAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACACTCC 929  
Qy 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300  
Db 930 TTTCAATACGCTTTCTGTGAGCGATGTTTAAATATGATGATCTCTTCAATCCTTTTCA 989  
Qy 301 AlaThrProAsnThrTyrLysArgLysAsnThrGlnThrAlaLeuAspAsnLysProCys 320  
Db 990 GCAACACCCCAACACTTATTAAGCGAGAGAAACAGAAACAGCTCTAGACAAACCTTGT 1049  
Qy 321 GlnProGlnCysTyrGlnHisLeuGlnGlnValLysGlnPheAlaAlaLeuThrAla 340  
Db 1050 GAGCACAGTGTTCACGACATTTTGAAGGAGCAAAAGAGTTTGTCTGCTCTCACCGCT 1109  
Qy 341 GlnArgIleLysThrProProLysArgProGlyLysArgArgArgGlyArgLeuProAsn 360  
Db 1110 GAGCGGATTAAGACCCCAACCAAAAGTCAGGAGGCGCAGAGAGAGAGAGAGAGAGAG 1169  
Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnLysLysArgThrAspSer 380  
Db 1170 AACAGTGCAGAGCCACACACCCCAACCAATTAATGTGTGGAATCAAGAGATACAGACGT 1229  
Qy 381 AspArgGlnAlaGlyThrGlnLysThrGlyGlnLysAsnAspLysGlnGlnLys 400

Db 1230 GATAGGAGGAGGAGCTGAAACGGGGGAGAGAACATGATTAAGAAAGAGAG 1289  
Qy 401 LysApGluThrSerSerSerSerGluAlaAsnSerArgGlnThrProIleuMet 420  
Db 1290 AAGATGAACTTGACGCTCTCTGAGCAAAATTCGCGTCTCAACCAATTAAGATG 1349  
Qy 421 LysProAsnIleGluProGluAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440  
Db 1350 AAGCAAAATATTGAACCTCGTAGAATGTGAGTGGATGGTGGAGCCCTCAATGTTT 1409  
Qy 441 ArgValIleuIleGlyThrTyrThrAspAsnMetCysAlaIleAlaArgLeuIleGlyThr 460  
Db 1410 AGAGTCCCTCATTTGGACTTACTATGACATTTCTGTGCATTTGCTAGGTATTTGGAGCC 1469  
Qy 461 LysThrCysArgGlnValTyrGluPheArgValLysGlnSerSerIleIleAlaProAla 480  
Db 1470 AAAACATGTAGACAGGTATGAGTTTGAAGTCMAAGATCTAGCATCATAGCTCCAGCT 1529  
Qy 481 ProAlaGluAspValAspThrProProArgLysLysArgLysAsnIleArgLeuTrpAla 500  
Db 1530 CCGCTGAGATGTGATGATCTCTCAAGAAAGAGAGAAACACCGGTGTGGGCT 1589  
Qy 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
Db 1590 GCACACTGCGAGAAAGATACAGCTGAAAAGACGGGCTCTTAAACATGTTTCAACTAT 1649  
Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
Db 1650 CAACCTGTGATCATTCACGGGAGGCTTGTGACAGTTGTCCTGTGTGTATGACACAA 1709  
Qy 541 AsnPheCysGluLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
Db 1710 AATTGTTTGAAAAGTTTGTCAATGTAGTTCAAGTGTCAAAACCGCTTCCGGAGTGC 1769  
Qy 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
Db 1770 CGCTGCAAGACCACTGTGAAACCAAGACGTCCGCTGCTGCTGCTGCTGCTGCTGCT 1829  
Qy 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTyrAspSerLysAsnValSer 600  
Db 1830 GACCTGACCTCTGTCTTACTTGTGAGCCGCTGACCAATGGAGACAGTAAATGTGTCC 1889  
Qy 601 CysLysAsnCysSerIleGlnArgGlySerLysLysIleLeuLeuAlaProSerAsp 620  
Db 1890 TGCAAGAACTGCGATTCAGCGGGGCTCCAAAAGACATCTATCTGTGCAACATCTAC 1949  
Qy 621 ValAlaGlyTTPGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
Db 1950 GTGGAGGCTGGGGGATTTTATCAAGATCTGTGCAAGAAATGAAATTCATCTCAAGAA 2009  
Qy 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgGlyLysValTyrAsnLys 660  
Db 2010 TACTGTGAGACATTTATTTCTCAACATGACGTGACAGAGGGAAGTGTATATATAA 2069  
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAlaAspAlaThrArgLys 680  
Db 2070 TACATGTGACAGCTTCTGTTCACCTTGAACATGATTTTGTGTGATGACAAACCGCAAG 2129  
Qy 691 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
Db 2130 GGTAACAAAATTCGTTTGGCAATCATTCGTGTAATCCAACTGCTATGCAAAAGTT 2186

FILE REFERENCE: 0652.1670001  
CURRENT APPLICATION NUMBER: US/10/773.302  
CURRENT FILING DATE: 2004-02-09  
PRIOR APPLICATION NUMBER: US/09/589,892  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: US 08/945,988  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: PCT/EP96/01818  
PRIOR FILING DATE: 1996-05-02  
PRIOR APPLICATION NUMBER: DE 195 16 776.7  
PRIOR FILING DATE: 1995-05-10  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(89)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (90)..(2330)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (2331)..(2600)  
US-10-773-302-1  
Alignment Scores:  
Pred. No.: 0 Length: 2600  
Score: 3808.00 Matches: 699  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0  
US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-773-302-1 (1-2600)  
Qy 1 MetGlyGlnThrArgLysLysSerGluLysGlyProValCysTyrPargLysArgValLys 20  
Db 90 ATGGGCGACAGCTGGGAAGAAATCTGAGAGGAGCAAGTTGTTGGCGAGCGGTGAAAA 149  
Qy 21 SerGluTyrMetArgLeuArgGlnLeuLysArgPheArgArgAlaAspGluValLysSer 40  
Db 150 TCAAGATGACAGCACTGAGACAGCTCAAGAGGTTCAAGACGCTGATGAAGTAAAGAT 209  
Qy 41 MetPheSerSerAsnArgGlnLysIleGluGluArgThrGluIleLeuAsnGlnGluTrp 60  
Db 210 ATGTTTATGTTCCATTCGTCAAGAAATTTTGAAGAAACGAAATCTTAAACCAAGATG 269  
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 270 AAAACGAGGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTTGCGGGACT 329  
Qy 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
Db 330 AGGAGGAGTTCGGGACCAAGTGAATTTTCCAAACAAAGCATCCATTAAAGAT 389  
Qy 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnAlaAsnPhe 120  
Db 390 CTGAATGCAGTGCTTCAAGTACCAATATGATTTCTGTGCTCCCTTACAGCAAAATTTT 449  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
Db 450 ATGGGGAAGATGAACGTTTTCATPAACATTCCTTATATGAGAGATGAAGTTTATGAT 509  
Qy 141 GlnAspGlyThrPheIleGluGlnLeuLysAsnTyrArgGlyLysValHisGlyAsp 160  
Db 510 CAGATGTGACTTTCATTTGAAGAACTAATATAAATAATATGATGGGAAAGTACAGGGGAT 569  
Qy 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 570 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGTTGTGATGCCCTTGTGTCA 629



QY 41 MetPheSerSerAsnArgGlnLysIleLeuGlnArgThrGlnIleLeuAsnGlnGlnIleTrp 60  
DB 287 ATGTTTAGTTCACATCGTCAGAAAATTTGGAAAAGAACGAAATCTTAAACCAAGAAATGG 346  
QY 61 LysGlnArgGlnIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
DB 347 AAACAGCGAAGAAATACAGCTGTGTGACACATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 406  
QY 81 ArgGlnCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
DB 407 AGGAGAGTTCGCGTGAACAGTGAAGTTCGATTTTCCACACAAAGTATCCATTAAAGACT 466  
QY 101 LeuAsnAlaValIleAspValProIleMetYrSerTrpSerProLeuGlnAspPhe 120  
DB 467 CTGATGAGATGCTTCAGTACACCATATGATTCCTGTCTCCCTCAGACAGAAATTTT 526  
QY 121 MetValGlnAspGlnThrValLeuHisAsnIleProYrMetGlyAspGlnValLeuAsp 140  
DB 527 ATGGTGAAGATGAAATGTTTACATACATCTCTATATGAGATGAAAGTTTATGAT 586  
QY 141 GlnAspGlyThrPheIleGlnGlnLeuIleLysAsnYrAspGlyLysValHisGlyAsp 160  
DB 587 CAGATGCTACTTTCACTTGAAGAACTAATAAAATTTATGATGGGAAAGTACACGGGAT 646  
QY 161 ArgGlnCysGlyPheIleAsnAspGlnIlePheValGlnLeuValAsnAlaLeuGlyGln 180  
DB 647 AGAGAAATGTGGTAAATGATGATGAAATTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 706  
QY 181 TyrAsnAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 200  
DB 707 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 766  
QY 201 LysAspLeuGlnAspHisAspAspAspLysGlnSerAspProProArgLysPheProSer 220  
DB 767 AAAGATCTGAGAGATCACCGAGATGATAAAGAAACCGCCACCTCGGAAATTTCTTCT 826  
QY 221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGlnGlnLeu 240  
DB 827 GATAAAATTTTGAAGCCATTTCCATGTTTCCAGATTAAGGGACACGACAGAAAGACTA 886  
QY 241 LysGlnLysYrLysGlnLeuThrGlnGlnLeuProGlyAlaLeuProProGlnCys 260  
DB 887 AAGGAAAATATTAAGAACTCACCGAACACGCTCCACGGGACCTTCTCTGAAATGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnIleSerLeuHisSer 280  
DB 947 ACCCCCAACATAGATGACCAAAATCTTAATCTGTTCAGAGAGCAAAAGCTTACACTC 1006  
QY 281 PheHisThrLeuPheCysArgArgCysPheLysYrAspCysPheLeuHisProPheHis 300  
DB 1007 TTTCAATCGCTTTCTGTAGGCGATGTTTAAATGTGACTGCTTCTCAATCCTTTTCAAT 1066  
QY 301 AlaThrProAsnThrYrLysArgLysAsnThrGlnIleuAlaLeuAspAsnLysProCys 320  
DB 1067 GCACACCCCAACATTAATAGCGAAGAACACACAAACAGCTTTCAGACAAACACTTGT 1126  
QY 321 GlyProGlnCysYrGlnHisIleLeuGlnGlyAlaLysGlnPheAlaAlaIleuThrAla 340  
DB 1127 GGAACACAGTGTATTCACAGATTTTGAAGGAGCAAAAGGATTTGCTGCTGCTTCAACCT 1186  
QY 341 GlnArgGlnIleLysThrProProLysArgProGlyGlyArgArgArgGlyValArgLeuProAsn 360  
DB 1187 GAGCGGATTAAGACCCCAACAAACGTCACAGAGGCCACACAGAAAGAGCGGCTTCCCAAT 1246  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
DB 1247 AACAGTACAGAGCCCAACGACCCCAACATTAATGTGTGGAATCAAAAGATACACACAGT 1306  
QY 381 AspArgGlnAlaGlyThrGlnThrGlyGlnLysGlnAsnAspLysGlnGlnGlnLys 400  
DB 1307 GATAGGGAAGCGAGGATGAAACGGGGGAGAGAAATGATGATTAAGAAAGAAAGAAAGAG 1366  
QY 401 LysAspGlnThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420

DB 1367 AAAGATAAACCTTCGACCTCTCGAAGCAAAATTCGTGCTCAACCAATTAAGATG 1426  
QY 421 LysProAsnIleGlnProProGlnAsnValGlnTrpSerGlyAlaGlnAlaSerMetPhe 440  
DB 1427 AAGCCAAATATTAAGCTCTCTGAGAAATGTGAGAGTGTGTGTGTGTGTGTGTGTGTGT 1486  
QY 441 ArgValLeuIleGlyThrYrYrAspAspPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1487 AGAGTCTCATTTGGCAGTCACTATGACAAATTTCTGTGCTCAATGCTAGATTAATTTGGACC 1546  
QY 461 LysThrCysArgGlnValYrGlnLysPheArgValLysGlnSerSerIleIleAlaProAla 480  
DB 1547 AAACATGTAGACAGGTGTATGAGTTAGAGTCAAAATCTTACATCATAGCTCCAGCT 1606  
QY 481 ProAlaGlnAspValAspThrProProArgLysLysLysValArgLysHisValArgLeuTrpAla 500  
DB 1607 CCGCTGAGAGATGTGATATCTCTCCAAAGAAAAGAAAGAAAGAACCCGGTTGTGGGCT 1666  
QY 501 AlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsnHisValYrAsnYr 520  
DB 1667 GCACACTGCAAGAAAGATACAGCTGAAAAGACCGCTCTCTAACCATGTTTACAACTAT 1726  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1727 CAACCTGTGATCATCCACCGCAGCTTGTGACAGTTGCGTCCCTGTGTGATGACAA 1786  
QY 541 AsnPheCysGlnLysPheCysGlnCysSerSerGlnCysGlnAsnArgPheProGlyCys 560  
DB 1787 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTGCAAAACCGCTTCCGGGATGC 1846  
QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrIleuAlaValArgGlnCys 580  
DB 1847 CGCTGCAAGCAAGTCAACACCAACAGAGCCCGTCAACCTGCTGTCCAGAGATGT 1906  
QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
DB 1907 GACCTTACCTCTGTCTTACTGTGTGAGCGCTGTACCATTTGGACAGTAAATGTGTCC 1966  
QY 601 CysLysAsnCysSerIleGlnArgLysSerLysLysIleLeuLeuAlaProSerAsp 620  
DB 1967 TGCAAGAACTGCATATTAAGCGGGGCTCCAAAAGCATGATGATTCGGACCATCTGAC 2026  
QY 621 ValAlaGlyTrpGlyIlePheIleLysAspProValGlnLysAsnGlnPheIleSerGln 640  
DB 2027 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAAAATGAATCATCTCAGAA 2086  
QY 641 TyrCysGlnGlnIleIleSerGlnAspGlnAlaAspArgArgGlyLysValYrAspLys 660  
DB 2087 TACTGTGAGAGATTAATTTCTCAAGATGAGCTACAGAAAGGGAAGGTATGATAAA 2146  
QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680  
DB 2147 TACATGTGAGAGCTTCTGTTCACATTCGAAATGATTTTGTGTGTGTGCAACCCGCAAG 2206  
QY 681 GlyAsnLysIleAspPheAlaAsnHisSerValAsnProAsnCysYrAlaLysVal 699  
DB 2207 GGTAAACAAATTCGTTTGGCAATCATTCGGTAAATCAAACTGCTAAGCAAAAGTT 2263

RESULT 5  
US-10-153-668-231  
; Sequence 231, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; TITLE OF INVENTION: STATE Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; PRIORITY FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172

PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: US 60/316,031  
PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: US 60/328,403  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: JP 2001-157043  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: JP 2001-260681  
PRIOR FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: JP 2001-313175  
NUMBER OF SEQ ID NOS: 488  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 231  
LENGTH: 2558  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (58)..(2295)  
US-10-153-668-231

Alignment Scores:  
Pred. No.: 0 Length: 2558  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.8% Indels: 0  
Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-153-668-231 (1-2558)

1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20  
58 ATGGCCAGACTGGGAGAAATCTGAGAAAGGAGCAGTTGTGGCGGAGCGTAAAA 117  
21 SerGluTrpMetArgLeuArgLysLysArgPheArgGlyAlaSerGluValLysSer 40  
118 TCAGAGTCAATGCACTGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAAGTAAAGT 177  
41 MetPheSerSerAspArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
178 ATGTTTACTTCCTCAATCTGCAAAATTTGGAAAGACGAAATCTTAAACCAAGATG 237  
61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
228 AAACGCGAAGATACAGCTGTGCAATCTGACTTCTGTGAGCTCATTCGCGGAGCT 297  
81 ArgGluCysSerValThrSerArgPheLeuAspPheProThrGlnValIleProLeuLysThr 100  
228 AGGAGGTTCGGTACCACTGATGATGATTTCCAAACAGATCATCCATTAAAGCT 357  
101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120  
358 CTGAATGCACTTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 417  
121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyLysArgValLeuAsp 140  
418 ATGGGGAAGATGAAGAACTGTTTACATTAATTCCTTAATATGAGAGATGAAGCTTTAAGT 477  
141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGluLysAsp 160  
478 CAGGATGGTACTTCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACT 537  
161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGluGln 180  
538 AAGAAATGGGTTTAAATATGAATATTTTGTGAGATTGGTGAATCCCTTGTCTCA 597  
538 AAGAAATGGGTTTAAATATGAATATTTTGTGAGATTGGTGAATCCCTTGTCTCA 597  
181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluLysGluGluLysGln 200  
598 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220

658 AAAGTCTGGAGATCAACGAGATGATTAAGAAAGCCGCCACCTCGGAAATTCCTCT 717  
221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
718 GATAAATATTTTGGAGGCGCATTTCTCAATGTTTCCAGTAAAGGCGACGAGAAAGCTA 777  
241 LysGluLysGlyLysGluLeuThrGlnGlnIleLeuProGluValAlaLeuProGluLys 260  
778 AAGAAATATTAAGAACTCAACGACAGACTCCAGGCGCATTCCTCTGAAATCT 837  
261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
838 ACCCCCAATATATGAGCAGAAATGCTTAATCTGTGAGAGAGAGCAAGCTTACATCC 897  
281 PheHisIlePhePheCysValArgCysPheLysGlyThrCysPheLeuHisIleProPheHis 300  
898 TTTCAATAGCTTTTCTGATGAGGATGATGATGATGATGATGATGATGATGATGAT 957  
301 AlaThrProAsnThrTyrLysArgLysValAsnThrGluThrAlaLeuAspAsnLysProCys 320  
958 GCAACACCCCACTTATTAAGCGGAGAAACACAGAAACAGCTTAGACAAACACCTTGT 1017  
321 GlyProGlnCysTyrGlnHisIleLeuGluGluValAlaLysGluPheAlaAlaAlaLeuThrAla 340  
1018 GACCAACAGTCTTACAGCATTTGGAGGAGCAAAAGAGTTTGTGCTGCTGCTGACGCT 1077  
341 GluArgIleLysThrProProLysArgProGluGluValArgArgGlyValGluLeuProAsn 360  
1078 GAGCGGATTAAGACCCCAAAACGTCACAGAGCCGACAGAGAGCGGCTTCCCAT 1137  
361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
1138 AACAGTACAGGCGCCAGACCCCAACATTAATGCTGGAATCAAGAGATACAGACAGT 1197  
381 AspArgGluAlaGlyThrGlnThrGluGluLysLeuAsnAspLysGluGluGluLys 400  
1198 GATAGGAGAGCAGGAGCTGAACCGGGGAGAGAAATGATTAAGAAAGAAAGAGAG 1257  
401 LysAspGluThrSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
1258 AAAGTGAATCTGAGCTCTGAGCAATTTCTCGGTGCAACACCAATTAAGATG 1317  
421 LysProAsnIleGluProProGluAsnValGluTrpSerGluValAlaGluAlaSerMetPhe 440  
1318 AAGCCAAATATTAAGCTCTGAGATGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1377  
441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
1378 AGATCTCTATTTGCACTTACTATGACAAATTTCTGTGCAATTTCTGATTAATGGAGCC 1437  
461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
1438 AAACATGTAGACAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1497  
481 ProAlaGluAspValAspThrProProArgLysLysValGluValHisArgLeuThrAla 500  
1498 CCGGTGAGATGGAGATACCTCTCCAGAGAAAGAAAGAGAGAAACACCGGTGTGGGCT 1557  
501 AlaHisCysArgLysIleGlnLeuLysValAspGlySerSerAsnHisValTyrAsnTyr 520  
1558 GCACTCAAGAAAGATACAGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
521 GlnProCysAspHisIleProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
1618 CAACCTGTATTCATCCAGCGAGCTGTGACAGTTGTGCTGCTGTGATATGACADA 1677  
541 AsnPheCysGlyLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGluCys 560  
1678 AATTTTGTGAAAGTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgLysCys 580

Db 1738 CGCTGCAAGCAGATGCAACCAAGCAGTCCCGTCTACCTGCTGTCGAGAGTGT 1797  
Qy 581 AASPpRoAspLeuCySLeuThrCySglYAlaAlaAspHisTPAAsPserLYsAsnValSer 600  
Db 1798 GACCTTACCTCTCTCTTACTTCTGAGACCCCTGACCACTTGGACAGTAAATGTCGCC 1857  
Qy 601 CyAlYsAsnCySserIleGlnArgLYserLYsLYsHisLeuLeuLeuAlaProSerThAsP 620  
Db 1858 TGCAGAAAGCTCAGATTCAGCGGGGCTCCAAAAAGCATCTATTGCTGGCACCACATCGAC 1917  
Qy 621 ValAlaGlyTPpGlyIlePheHileYsAsPProValGlnLYsAsnGlnPheIleSerGlu 640  
Db 1918 GTGGCAGGCTGGGGATTTTATCAAAAGATCTGTGCGAAAAATGATTCATCTCGAA 1977  
Qy 641 TyrCySglYgluIleIleSerGlnAspGluAlaAspArgArgLYsValTyrAspLYs 660  
Db 1978 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAAAGGAAAGTATGATTA 2037  
Qy 661 TyrMetCySserPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLYs 680  
Db 2038 TACATGTGCACCTTTCTGTCAACTTGAAACATGATTTTGTGTGATGCAACCGCAG 2097  
Qy 681 GLYsAsnLYsIleArgPheAlaAsnHisSerValAsnProAsnCySTyrAlaLYsVal 699  
Db 2098 GGTAAACAATAATCGTTTGGCAATCATTCGGTAAATCCAAACTGCTATGCAAAAGTT 2154

## RESULT 6

us-10-172-118-1043  
Sequence 1043, Application US/10172118  
Publication No. US20030224374A1  
GENERAL INFORMATION:  
APPLICANT: Dai, Hongyue  
APPLICANT: He, Yudong  
APPLICANT: Linaley, Peter  
APPLICANT: Mao, Mao  
APPLICANT: Roberts, Chris  
APPLICANT: Van 't Veer, Laura  
APPLICANT: Van de Vijver, Marc  
APPLICANT: Bernards, Rene  
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
FILE REFERENCE: 9301-175-899  
CURRENT APPLICATION NUMBER: US/10/172,118  
CURRENT FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: 60/380,770  
PRIOR FILING DATE: 2002-05-14  
NUMBER OF SEQ ID NOS: 2699  
SEQ ID NO 1043  
LENGTH: 2576  
TYPE: DNA  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
DATABASE ACCESSION NUMBER: NM\_004456  
DATABASE ENTRY DATE: 2001-06-18  
us-10-172-118-1043

## Alignment Scores:

Pred. No.: 0 Length: 2576  
Score: 3802.00 Matches: 698  
Percent Similarity: 99.9% Conservative: 0  
Best Local Similarity: 99.9% Mismatches: 1  
Query Match: 99.8% Indels: 0  
DB: 7 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-172-118-1043 (1-2576)

Qy 1 MetGlyGlnThrGlyLYsLYsSerGluYsGlyProValCySTyrArgLYsArgValLYs 20  
Db 58 ATGGGACAGACTGGGAGAAATCTGAGAAAGGAGCAGGTTTGTGGCGGAGCGTGTAA 117  
Qy 21 SerGlyTyrMetArgLeuArgGlnLeuLYsArgPheArgArgAlaAspGluValLYsSer 40  
Db 118 TCAGAGTACATGCGACTGAGCAGCTCAAGAGGTTCAAGCAGCTGATGAGATTAAGAGT 177

Qy 41 MetPheSerSerAsnArgGlnLYsIleLeuGluArgThrGlnIleLeuAsnGlnIleTyr 60  
Db 178 ATGTTGTGTTCCCATCTCGAGAAATTTTGGAAAGACGAAATCTTAAACCAAGAAATGG 237  
Qy 61 LYsGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
Db 238 AACACGCGAAGAGATACGCTGTGACACTCTGACTTCTGTGAGCTCATTTGGCGGGAGCT 297  
Qy 81 ArgLYsCySserValThrSerAspLeuAspPheProThrGlnValIleProLeuLYsThr 100  
Db 298 AGGAGAGTTCGGGACCACTGACTTGATTTTCCAAACAGATCATCCCATTAAGAACT 357  
Qy 101 LeuAsnAlaValAlaSerValProIleMetTyrSerThrSerProLeuGlnIleAsnPhe 120  
Db 358 CTGAATCAGTGTGTTCAAGTACCATATGATTTCTTGCTCCCTACAGCAGAAATTTT 417  
Qy 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
Db 418 ATGTTGGAAGAGTAACTGTTTACATACATTTCTTATATGGAGATGAAGTTTATGAT 477  
Qy 141 GlnAspGlyThrPheIleGlnGluLeuIleLYsAsnTyrAspGlyLYsValHisGlyAsp 160  
Db 478 CAGAGTGTACTTTCATTTGAAGAACTAATAAATATATATGAGGAAAGTACACGGGGAT 537  
Qy 161 ArgGluCySglYpHeIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
Db 538 AGAGAAATGTGGGTTTATTAATGATGAATTTGTGTGAGTTGTGAATGCCCTTGTGTC 597  
Qy 181 TyrAsnAspAspAspAspAspAspAspLYsAspAspProGluGluArgGluGluLYsGln 200  
Db 598 TATTAATATGATGACATGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAACAG 657  
Qy 201 LYsAspLeuGluAspHisArgAspAspLYsGluSerArgProProArgLYsPheProSer 220  
Db 658 AAAGATCTGGAGGATCAACGAGATGATTAAGAAAGCGCCACCTCGGAAATTTCTTCT 717  
Qy 221 AspLYsIlePheGlnAlaIleSerSerMetPheProAspLYsGlyThrAlaGluGluLeu 240  
Db 718 GATMAAATTTTGGGGCATTTTCTCAATGTTTCCAAATTAAGGACACAGCAAGAAACTA 777  
Qy 241 LYsGluLYsTyrLYsGluLeuThrGluGlnLeuProGluAlaLeuProProGluLYs 260  
Db 778 AAGAAATAATTAAGAACTCACCAAGACAGCTCCACAGCGCATCTTCTCTGATGT 837  
Qy 261 ThrProAsnIleAspGlyProAsnAlaLYsSerValGlnArgGluGlnSerLeuHisSer 280  
Db 838 ACCCCCAACATAGATGAGCCAAATGCTAATCTGTTCAGAGAGCAAAAGCTTACACTCC 897  
Qy 281 PheHisThrLeuPheCySArgArgCySArgPheLYsTyrAspCySArgPheHisArgProPheHis 300  
Db 898 TTTCAATGCGCTTTTCTGTAGCGATGTTTAAATATGATGATGCTTCTTACATCTTTTCAT 957  
Qy 301 AlaThrProAsnThrTyrLYsArgLYsAsnThrGlnThrAlaLeuAspAsnLYsProCYs 320  
Db 958 GCAACACCCACACTTATTAACCGAAGAACACAGAAACAGCTCTAGACACAAACCTTGT 1017  
Qy 321 GLYProGlnCYsTyrGlnHisIleLeuGluGlyAlaLYsGluPheAlaAlaIleuThrAla 340  
Db 1018 GAGCACAAGTGTTCACGACATTTGAGAGGACAAAGAGTGTGCTGCTCACCGCT 1077  
Qy 341 GlnArgGlyLeuThrProProLYsArgProGlyGlyArgArgArgGlyArgGluLeuProAsn 360  
Db 1078 GAGCGGATTAAGACCCCAACAAAGCTCCAGAGAGCGCAGAAAGAGAGCGGCTTCCCAT 1137  
Qy 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLYsAspThrAspSer 380  
Db 1138 AACAGTGCAGGCCCCAGACCCCACTTATATGTCTGGAATCAAGAGATACAGACAT 1197  
Qy 381 AspArgGluAlaGlyThrGlnGluThrGlyGlyGluAsnAspLYsGluGluGluGluLYs 400  
Db 1198 GATAGGAGAGCAGGAGCTGAAGACGGGGGAGAGAAACATGATTAAGAAAGAGAGAGAG 1257  
Qy 401 LYsAspGluThrSerSerSerSerGluAlaAsnSerArgCySglnThrProIleLYsMet 420

```

Db 1258 AAAGTGAAGAACTTCAGAGCTCCTCTGAGAGAAATTCCTGGTGCACAAACCATTAAGATG 1317
Qy 421 LysProAsnIleGluProGluAsnValIgluTPSerGlyValGluAlaSerMetPhe 440
Db 1318 AAGCCAAATATTAACCTCTCTGAGATGAGAGTGGAGTGGTGCAGACCTCAATGTTT 1377
Qy 441 ArgValIleuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgIleuIleGlyThr 460
Db 1378 AGAGTCCATGAGGACCTTACTATATGACAAATTCCTGGCTGATGCTTAATTTGGAGCC 1437
Qy 461 LysThrCysArgGlnValTyrGluPheArgValIlyGluSerSerIleIleAlaProAla 480
Db 1438 AAAACATGTAAGACAGGTGATAGATTAGCTCAAGAAATCTAGCATCACTACCTCAGCT 1497
Qy 481 ProAlaGluAspValAspThrProArgIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 500
Db 1498 CCGGCTGAGAGATGGATGACTCCTCCAAAGAAAAGAAAGAGAAACACCGGTTGTGGGCT 1557
Qy 501 AlaHisCysArgIlyIleGluIleuIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 520
Db 1558 GCACACTGCAAGAAATACAGCTGAAAAGAGACGGCTCCTTACCAATGTTTACAACTAT 1617
Qy 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
Db 1618 CAACCTGTGATCATCCACGAGGCTTGTGACAGTTGCTGCTGTGTGTATGACACA 1677
Qy 541 AsnPheCysGluIlyPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560
Db 1678 AATTTTGTGAAAAGTTTGTGCAATGTAGTTCAAGTGTCAAAACCGCTTCCGGAGATGC 1737
Qy 561 ArgCysIlyIleAlaGlnCysAsnThrIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 580
Db 1738 CGCTGCACCAAGCAGCTGCAACCAAGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1797
Qy 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerIlyAsnValSer 600
Db 1798 GACCTGTGACCTTGTCTTACTTGTGAGCCGCTGACCAATGGACAGTAATAATTTGTC 1857
Qy 601 CysIlyAsnCysSerIleGlnArgGlySerIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 620
Db 1858 TGCAAGAACTGCAATTCAGCGGGGCTCCAAAAGCAATCTATCTGCGCACATCTGAC 1917
Qy 621 ValAlaGlyTTPGlyIlePheIleIlyAspProValGlnIlyAsnGluPheIleSerGlu 640
Db 1918 GTGGAGGCTGGGGGATTTTATCAAGATCCTGTGCAGAAAATGAATTCATCTCAGAA 1977
Qy 641 TyrCysGlyIlyIleIleSerGlnAspGluIleAspArgGlyIlyValIlyIlyIlyIly 660
Db 1978 TACTGTGAGAGATTTATTTCTCAAGATGAGCTGACAGAGGGGAAAGTGTATATAA 2037
Qy 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValIleAspAlaThrArgIly 680
Db 2038 TACATGTGCAAGCTTTCTGTCAACTTGAACAATGATTTTGTGTGTGAGTCAACCGGAG 2097
Qy 681 GlysAsnIlyIleArgPheAlaAsnHisSerValIleProAsnCysIlyIlyIlyVal 699
Db 2098 GGTAACAAATTCGTTTGGCAATCATTCGTAATTCAAACCTGATGCAAAAGTT 2154

```

# RESULT 7

```

US-10-295-027-709
; Sequence 709, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezil, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Waterson, Susan R.

```

```

; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295, 027
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/340, 376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 709
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-709

Alignment Scores:
Pred. No.: 0 Length: 2576
Score: 3802.00 Matches: 698
Percent Similarity: 99.9% Conservative: 0
Best Local Similarity: 99.9% Mismatches: 1
Query Match: 99.8% Indels: 0
DB: 7 Gaps: 0

US-10-773-302-2_COPY_1_699 (1-699) x US-10-295-027-709 (1-2576)
Qy 1 MetGlyInthrGlyIlyIlySerGluIlyGlyProValCysTrpArgIlyArgValIly 20
Db 58 ATGGGCCAGACTGGGAGAAATCTGAGAGGAGCAAGTTGTTGGCGGAAACGTAAATA 117
Qy 21 SerGluTyrMetArgLeuArgGlnLeuIlyArgPheArgArgAlaAspGluValIlySer 40
Db 118 TCAGAGTACAGACCACTGACACAGCTCAAGAGTTTCAGACGCTGATGAAGTAAAGAT 177
Qy 41 MetPheSerSerAsnArgGlnIlyIleGluIlyArgThrGluIlyIleAsnGluIlyTrp 60
Db 178 ATGTTGTTTCATTCGTCAGAAAATTTTGGAAAAGAACGAAATCTTAACAGAAATGG 237
Qy 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValIleSerSerLeuArgGlyThr 80
Db 238 AAACAGGAGAGATACAGCTGACATCTGACTTCTGTGACTCATTTGCGGGAGACT 297
Qy 81 ArgGluCysSerValIlyThrSerAspLeuAspPheProThrGlnValIleProLeuIlyThr 100
Db 298 AGGAGTGTTCGGGAGCAAGTGTGATTTTCCAAACAGATCATCCCAATTAAAGACT 357
Qy 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnIleAsnPhe 120
Db 358 CTGAATGAGTGTCTTCAATGACCAATATGATTTCTGTCTCCCTTACAGCAAAATTTT 417
Qy 121 MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyIlyAspGluValIleuAsp 140
Db 418 ATGTGTGAGATGAAACGTGTTTATCATTAATCTTCTTATATAGGAGATGAAGTTTATGAT 477

```







## RESULT 9

US-10-848-755A-129  
; Sequence 129, Application US/10848755A  
; Publication No. US20050054826A1  
; GENERAL INFORMATION:  
; APPLICANT: Mao, Mao  
; TITLE OF INVENTION: HUMAN DIAPHANOUS-3 GENE AND METHODS OF USE THEREFOR  
; FILE REFERENCE: 9301-196-999  
; CURRENT APPLICATION NUMBER: US/10/848, 755A  
; PRIOR FILING DATE: 2004-05-18  
; PRIOR FILING DATE: 2003-05-11  
; NUMBER OF SEQ ID NOS: 275  
; SOFTWARE: PatentIn version 3.2 CAM: 301891-999188  
; SEQ ID NO 129  
; LENGTH: 2576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-848-755A-129

## Alignment Scores:

Pred. No.:	0	Length:	2576
Score:	3802.00	Matches:	698
Percent Similarity:	99.94	Conservative:	0
Best Local Similarity:	99.94	Mismatches:	1
Query Match:	99.84	Indels:	0
DB:	10	Gaps:	0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-10-848-755A-129 (1-2576)

QY 1 MetGlyGlnThrGlyLeuValSerGlyLeuValProValCysTrpArgLeuValGlyLeu 20  
DB 58 ATGGGCCAGACTGGGAGAAATCTGAGAAAGGACCATTTGTGGCGGAAACGGTATAA 117  
QY 21 SerGlyTyrMetArgLeuArgGlnLeuValArgPheArgGlyLeuValGlyLeuSer 40  
DB 118 TCAGAGTACATGCGACTGAGACAGCTCAAGAGTTCAGACGAGTGAAGTAAAGGT 177  
QY 41 MetPheSerSerAspArgGlnLeuValLeuGluArgThrGlnLeuAsnGlnGluTrp 60  
DB 178 ATGTTAGTTCACATCGTCAGAAATTTGGAAAGACGAAATTTTAAACCAAGATGG 237  
QY 61 LysGlnArgArgGlnProValHisLeuThrSerValSerSerLeuArgGlyLeu 80  
DB 238 AAACGCGAAGGATACAGCTGTGCACTCTGACTTCTGTGAGCTCATTTGGCGGAGCT 297  
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuValThr 100  
DB 298 AGGGAGTGTTCGGTACCAAGTGAATTTTCCAAACAGATCATCCCATTAAGACT 357  
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnAsnPhe 120  
DB 358 CTGAATGCGATTGCTTCAGTACCATTAATGTTCTTGGTCTCCCTTACAGCAAAATTT 417  
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140  
DB 418 ATGGTGAAGATGAATACGTTTATCATTAACATTTCTTATATGGAGATGAAGTTTATGAT 477  
QY 141 GlnAspGlyThrPheIleGluLeuIleValAsnTyrAspGlyLeuValHisGlyLeuAsp 160  
DB 478 CAGGATGATCTTATGAAAGACTAAATTAATAATTTGATGGAGAAAGTACACGGGAT 537  
QY 161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB 538 AGCAATGTGGGTTTATTAATGATGAATTTTGTGGAGTTGTGAATCCCTTGTGTCAA 597  
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluValGln 200  
DB 598 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 201 LysAspLeuGluAspHisArgAspAspLeuValGluSerArgProProArgGlyAspPheProSer 220  
DB 658 AAAGATCTGGAGGATCACCGAAGATGATTAAGAAAGCCGCCCATCTGGAAATTTCTTCT 717

QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspArgGlyThrAlaGluGluLeu 240  
DB 718 GATTAATTTTGGAGGCGCATTTCTCTCATGTTTCCAGATTAAGAGGACACAGCAAGAACTA 777  
QY 241 LysGluValSerTyrGluLeuThrGlnGlnIleLeuProGlyValLeuProProGluCys 260  
DB 778 AAGGAAAAATTAAGAACTACCGAAGACAGAGCTCCAGCGCATTTCTCTGAAATGT 837  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
DB 838 ACCCCCAACATATGATGACCAATGCTTAATCTGTTTCAAGAGAGCAAACTTACATCC 897  
QY 281 PheHisThrLeuPheCysArgArgCysPheLeuValTyrAspCysPheLeuHisProPheHis 300  
DB 898 TTTCATAGCTTTTCTGTAGCGCATGTTTAAATATGATGATGATGATGATGATGATGATGAT 957  
QY 301 AlaThrProAsnThrTyrValArgValAsnThrGluThrAlaLeuAspAsnValProCys 320  
DB 958 GCACACCCCAACACTTAAAGCGGAGAGAACACAGAAACAGCTTACGACAAACCTTGT 1017  
QY 321 GlyProGlnCysTyrGlnHisLeuGluGlyValAlaValGluPheAlaAlaLeuThrAla 340  
DB 1018 GACCAACAGTGTTCACAGCATTTGAGAGGACAAAGAGTTGTGCTGCTTACACCGCT 1077  
QY 341 GluArgIleLeuThrProProLysArgProGlyGlyValArgArgGlyValArgLeuProAsn 360  
DB 1078 GAGCGGATTAAGACCCCAACAAACCTCCAGAGGCGCGAGAAAGAGCGCTTCCCAAT 1137  
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLeuAspThrAspSer 380  
DB 1138 AACAGTACAGGCGCCACACCCCAACATTAATGTCTGAAATCAAGATACAGACAT 1197  
QY 381 AspArgGluAlaGlyThrGlnThrGlyGlyValAsnAsnAspLysGluGluGluVal 400  
DB 1198 GATGGAGAGCAGGAGCATGAACCGGGGAGAGAGAACATGATTAAGAAAGAAAGAGAG 1257  
QY 401 LysAspGluThrSerSerSerSerSerGlnAlaAsnSerArgCysGlnThrProIleLysMet 420  
DB 1258 AAAGATGAACCTTGACCTCCTGAAAGCAAAATCTCGGTCTCAAAACCAATTAAGATG 1317  
QY 421 LysProAsnIleGluProProGluValAsnValGluThrSerGlyAlaGluAlaSerMetPhe 440  
DB 1318 AAGCAATATATGAACCTCCTGAGAAATGTGAGAGTGTGTGAAAGCTCAATGTTT 1377  
QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
DB 1378 AGATCTCTCATTTGCACTTACTATGACAAATTTCTGTGCCANTGTAGGTTAATGGGACC 1437  
QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
DB 1438 AAAACATGTAACAGGTATGAGTTTAAAGTCAAAAGATCTACATCATATGCTCCAGCT 1497  
QY 481 ProAlaGluAspValAspThrProProArgValLysValAspGlyValHisArgLeuThrAla 500  
DB 1498 CCCCCTGAGATGTGAGTACTCTTCAAGGAAAGAGAGAAACACGGGTGTGGCT 1557  
QY 501 AlaHisCysArgLysIleGlnLeuLysAspGlySerSerAsnHisValTyrAsnTyr 520  
DB 1558 GCACCTGCGAAGAAAGATACAGCTGAAGAAAGGCGGCTCTTAAACATGTTTCAACATAT 1617  
QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
DB 1618 CAACCTGTGATCATCCACGGCACCTGTGTACAGTTGTGCCCTTGTGTATGACCA 1677  
QY 541 AsnPheCysGlyLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
DB 1678 AATTTTGTGAAAGTTTGTCAATGTATGATTCAGAGTCTCAAAACCGCTTCCGGGATGC 1737  
QY 561 ArgCysValAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValArgGluCys 580  
DB 1738 CGCTGCAAGACAGTGAACCAACAGAGTGGCCGGTGTACTCTGGGTGTGAGAGTGT 1797

```

QY 581 AAPPProAspLeuCySeLeuThrCySeGlyAlaAlaAspHisTrpAspSerLeuValSer 600
DB 1798 GACCTGACCTCTGCTTCTTCTGAGAGCGGCTGACCATTTGGAGACGTAATAATGTGTCC 1857
QY 601 CySeLeuAsnCySeSerIleGlnArgGlySerIleuValSerHisLeuLeuValAspSerAsp 620
DB 1858 TCAGAGAACTGACAGTATTCAGCGGGGCTCCAAAAGCATCTATTGTGTCGACCATCTGAC 1917
QY 621 ValAlaGlyTrpGlyIlePheIleLeuAspProValGlnLeuAsnGlnPheIleSerGlu 640
DB 1918 GTGGAGGCTGGGGGATTTTATCAAAAGATCTGTGGCAAAAAGAAATTCATCTCAGAA 1977
QY 641 TyrCySeGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLeuValTyrAspLys 660
DB 1978 TACTGTGAGAGATTTTCTTCAAGATGAGCTGACAGAGGAGAAAGTGTATGATTA 2037
QY 661 TyrMetCySeSerPheLeuPheAsnLeuAsnAspPheValValAspAlaThrArgLys 680
DB 2038 TACATGTGACGCTTCTGTTCACATTTGAACATGATTTTGTGTGGATGCAACCGCAG 2097
QY 681 GlyAsnLeuValLeuArgPheAlaAsnHisSerValAsnProAsnCySeTyrAlaLysVal 699
DB 2098 GGTAACAATAATCGTTTGGCAATCATCGTAAATCCAACTGCTATGCAAAAAGTT 2154

RESULT 10
US-10-504-173-78
; Sequence 78, Application US/10504173
; Publication No. US20050202428A1
; GENERAL INFORMATION:
; APPLICANT: Akordia Limited
; TITLE OF INVENTION: Pluripotential Stem Cells
; FILE REFERENCE: PI01863W0
; CURRENT APPLICATION NUMBER: US/10/504,173
; CURRENT FILING DATE: 2004-08-11
; PRIOR APPLICATION NUMBER: 0203359.5
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 78
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-504-173-78

Alignment Scores:
Pred. No.: 0 Length: 2576
Score: 3802.00 Matches: 698
Percent Similarity: 99.9% Conservative: 0
Best Local Similarity: 99.9% Mismatches: 1
Query Match: 99.8% Indels: 0
DB: 10 Gaps: 0

US-10-773-302-2_COPY_1_699 (1-699) x US-10-504-173-78 (1-2576)

```

```

QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnGlnAsnPhe 120
DB 358 CTGAATGCAAGTGTCTTCAGTACCCCAATAGATTTCTTGCTCTCCCTACAGCAAAATTTT 417
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140
DB 418 ATGTGTGAAGATGAAGACGTGTTTACATTAACATCTCTTATATGGAGATGAAGTTTAGAT 477
QY 141 GlnAspGlyThrPheIleGlnGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160
DB 478 CAGGATGCTACTTTCATTAAGAACTAAATAAAATTTATGATGGCAAAAGTACACCGGAGT 537
QY 161 ArgGluCySeGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlnGln 180
DB 538 AGAGATATGTGGTTTATTAATGATGAATTTTGTGGAGTTGGTAAATGCTTGGTCA 597
QY 181 TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLysGln 200
DB 598 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
QY 201 LysAspLeuGluAspHisArgAspAspLysGlySerArgProProArgLysPheProSer 220
DB 658 AAAGATCTGGAGGATCACCGAGATGATTAAGAAAGCCGCCCACTCGGAAATTTCTTCT 717
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240
DB 718 GATTAATAATTTGGAGGCCATTTCTCAATGTTTCCAGATTAAGGGACACAGCAAAACATA 777
QY 241 LysGluLysTyrLysGluLeuThrGlnGlnGlnLeuProGluAlaLeuProGluLys 260
DB 778 AAGGAAATATTAAGAACTACCGAACAGACGCTCCAGGGGCACTTCTCTGATGT 837
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280
DB 838 ACCCCCAATGATGACCAAAATGCTAAATCTGTTGAGAGAGCAAAAGCTTACACTCC 897
QY 281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis 300
DB 898 TTTCATAGGCTTTTCTGTAGCGAGATGTTTAAATGATGCTCTTCAATCTTTTCT 957
QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320
DB 958 GCAACACCAACATTAATTAAGCGAAGAACACAGAAACAGCTTGAACAAACCTTGT 1017
QY 321 GlyProGlnCysTyrGlnHisLeuGlnGlyAlaLysGluPheAlaAlaLeuThrAla 340
DB 1018 GACACACAGTGTACACAGCATTTGGAGGAGCAAAAGAGTTGCTGCTCTCACCGCT 1077
QY 341 GlnArgIleLysThrProProLysArgProGluGlyValArgArgGlyValArgLeuProAsn 360
DB 1078 GAGCGGATTAAGACCCCAACAAACGTCACAGAGCGCCAGAAAGAGCGCTTCCCAT 1137
QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380
DB 1138 AACAGTAGACGCCCAAGACCCCAACATTAATGTGCTGGAATCAAGATACAGACAGT 1197
QY 381 AspArgGluAlaGlyThrGlnThrGlyGlyGlnAsnAspLysGluGlnGluLys 400
DB 1198 GATAGGAGAGGACGAGCATGAACGCGGGGAGAGAAACATGATTAAGAGAGAGAGAG 1257
QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420
DB 1258 AAAGATGAACCTTGAGCTCTCTGAAACAAATTCCTGCTGTCACCAATTAAGATG 1317
QY 421 LysProAsnIleGluProProGluLysValGluThrSerGlyAlaGluAlaSerMetPhe 440
DB 1318 AAGCCAAATATTAACCTCTGAGAAATGAGAGTGAAGTGTGCTGAAGCTCAATGT 1377
QY 441 ArgValLeuIleGlyLysThrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460
DB 1378 AGAGTCTCATTTGACATTAATGACAAATTTCTGTGCAATGTGATGATTAATGGAGC 1437

```

Qy	461	Ly5ThrCy5aTgInVal1YTrGInPheArgVal1ySglUserSer1le1le1a1ProCa1	480
Db	1438	AAAACATGTAGACAGGTGTATGAGTTTAAAGTCAAAAGATCTTACATCATAGCTCCAGCT	1497
Qy	481	Pro1aGInuAPbValAPbThProProArgrgYsYsYsArGLySH1aArGLyTPPa1a	500
Db	1498	CCCCCTGAGATGGATCTCTCTCCAGAGAAAAGAGAGAAACACCGGTTGTGGGCT	1557
Qy	501	Ala1a1Cy5aTgLy51leGInLeuLy5aPbGlySerSerAaNH1aVal1YrAaNH1Yr	520
Db	1558	GCAACACTGCAGAAAGATACAGCTGAAAAGGACGGCTCTCTAACATAGTTTACAATAT	1617
Qy	521	GInProCy5aAPbH1aProArgrGInProCy5aAPbSerSerCy5ProCy5Val1le1a1GIn	540
Db	1618	CAACCTGTGTATCTCCACCGGACGCTTGTACAGTTGTGTCCTTGTGTATGACAA	1677
Qy	541	AaNH1eCy5aGlyuYrAPbPheCy5aGInCy5SerSerGInuCy5aGInaAaArpPheProG1Cy5	560
Db	1678	AAATTTTGTGAAAGTTTGTGCATGTATGTCTCAAGTGTCAAAACCGCTTTCGGGATGC	1737
Qy	561	ArgCy5aLy5a1aGInCy5aAaNH1ThrLy5aGInuCy5ProCy5YrLeu1a1aValArGLuCy5	580
Db	1738	CGCTGCAGAAAGACAGTGCAGACCAAGACAGTCCCGTGTCTACCTGGCTGTCCAGAGGT	1797
Qy	581	AaPProAPbLeuCy5LeuThCy5aGly1a1a1aAPbH1a1rAPbSerLy5aAVal1Ser	600
Db	1798	GACCTTGACCTCTGTCTTACTTGTGGAGCCGCTGACCATTTGGACAGTAAAAATGTGTCC	1857
Qy	601	Cy5aLy5aAaNH1Cy5Ser1leGInaArGLySerLy5aSH1aLeuLeuA1aProSerAPb	620
Db	1858	TGCAGAACTCAGATATTCACGCGGGCTCCAAAAGCAATCATGTGTGGACACATCTGAC	1917
Qy	621	Val1a1aG1YrPGLy1lePheH1eLy5aAPProValGInuLy5aAaGInuPheH1eSerGInu	640
Db	1918	GTGGCAGGCTGGGGATTTTATTCAAAATCTCTGTGCAGAAAATGAATTCATCTCAGAA	1977
Qy	641	TyrCy5aGlyGInu1le1leSerGInaAPGInu1aAPaArgrGLy5aVal1YrAPbLy5	660
Db	1978	TACTGTGAGAGATATTTCTCAAGATGACCTGCACAAAGGGAAGTGTATGATTA	2037
Qy	661	TyrTherCy5aArpPheLeuPheAaNH1aAaNH1aAPbPheVal1aAPa1aThrArpLy5	680
Db	2038	TACATGTGCAGCTTCTGTCTTCAACTGAACATGATTTGTGGTGGATGCAACCGCAG	2097
Qy	681	G1yAaNH1yS1leArpPheAlaAaNH1aSerValaAProAaNH1YrAla1Ly5aVal	699
Db	2098	GCTAACAAATTCGTTCTTGCAAAATCATTCGCTAAATCCAAATGCTATGCAAAAGTT	2154
RESULT 11			
US-10-960-414-396			
; Sequence 396, Application US/10960414			
; Publication No. US20060074565A1			
GENERAL INFORMATION:			
; APPLICANT: MITLER, LANCE D.			
; APPLICANT: GEORGE, JOSH			
; APPLICANT: VEGA, VINSINSIUS B.			
; TITLE OF INVENTION: METHODS, SYSTEMS, AND COMPOSITIONS FOR CLASSIFICATION,			
; TITLE OF INVENTION: PROGNOSIS, AND DIAGNOSIS OF CANCERS			
; FILE REFERENCE: 38271-76067			
; CURRENT APPLICATION NUMBER: US/10/960,414			
; CURRENT FILING DATE: 2004-10-06			
; NUMBER OF SEQ ID NOS: 500			
; SOFTWARE: PatentIn version 3.3			
; SEQ ID NO 396			
; LENGTH: 2576			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-960-414-396			
Alignment Scores:			
Pred. No.: 0			
Score: 3802.00			
Percent Similarity: 99.9%			
Length: 2576			
Matches: 698			
Conservative: 0			

Best Local Similarity:	99.9%	Mismatches:	1
Query Match:	99.8%	Indels:	0
DB:	12	Gaps:	0
US-10-773-302-2_COPY_1_699 (1-699) x US-10-960-414-396 (1-2576)			
QY	1 MetGlyglnThrGlyLysLysSerGlnLysGlyProValCysTrpArgLysArgValLys	20	
DB	58 ATGGGCGGAGCTGGGAGAAATCTGAGAGGAGCAAGTTGTGTGGAGCCGGTTAAA	117	
QY	21 SerGlnTrpMetArgLeuArgGlnLeuLysArgPheArgArgAlaArgGluValLysSer	40	
DB	118 TCAAGGTACATGCCAGCTGAGCAGCTCAAGAGGTTCCAGCAGGCTGATGAAGTAAAGGT	177	
QY	41 MetPheSerSerAsnArgGlnLysLeuLeuGlnArgThrGluIleLeuAsnGlnGluTrp	60	
DB	178 ATGTTTACTTCCAACTCGTCAGAAAATTTTGGAAAAGAACGGAATCTTAAACCAAGAAATGG	237	
QY	61 LysGlnArgArgGlnLeuGlnProValHisIleLeuHisSerValSerSerLeuArgGlyThr	80	
DB	238 AAACGCGCAAGGATACAGCCTGTGTGACACTCTGACTTCTGTAGCTCATATGCGGGAGCT	297	
QY	81 ArgGluCysSerValThrSerAspLeuArgPheProThrGlnValIleProLeuLysThr	100	
DB	258 AGGAGGTGTTGGGTGACCAAGTACGTTGATTTTCCACACAAAGTCATCCATTAAAGACT	357	
QY	101 LeuAsnAlaValAlaSerValProIleMetCysTrpSerProLeuGlnAsnPhe	120	
DB	358 CTGATGTGAGTTGCTTCAGTACCCATATGTAATCTTGGTCTCCCTACAGCAAAATTTT	417	
QY	121 MetValGluAspGlnThrValLeuHisAsnIleProCysMetGlyAspGluValLeuAsp	140	
DB	418 ATGGGGAAGATGAAACGTGTTTACATVAACTCTCTTATATGGAAGATGAAGTTTATGAT	477	
QY	141 GlnAspGlyThrPheIleGluGlnLeuIleLysAsnTrpAspGlyLysValHisGlyAsp	160	
DB	478 CAGAGTGTACTTTCATTTAGGAAGACTATATAAAATTTATGATGGGAAGTACACGGGAGT	537	
QY	161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln	180	
DB	538 AGAGATGTGGGTATTATAATGATGAATTTTGTGGAGTTGGTGAAGCCCTTGTCGCA	597	
QY	181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnGluArgGlnLysGln	200	
DB	599 TATATATGATGATGAGATGATGATGATGAGACGATCTCTGAAGAAAGAAAGAAAGCAG	657	
QY	201 LysAspLeuGlnAspHisArgAspAspLysGlnSerArgProProArgLysPheProSer	220	
DB	659 AAAGATCTGGAGATCACCGAGATGATTAAGAAAGCCGCCACCTCGGAATTTCTCTCT	717	
QY	221 AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGlnLeu	240	
DB	718 GATTAATAATTTTGAAGGCATTTTCTCATATGTTTCCAGATTAAGGGCAGCAGAAAGACTA	777	
QY	241 LysGlnLysTyrLysGlnLeuThrGlnGlnGlnLeuProGlyValAlaLeuProGlnCys	260	
DB	778 AAGGAAAATATAAAGAACTCACGCAACAGAGCTCCAGCGGCACTTCTCCGTAATGT	837	
QY	261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGlnGlnSerLeuHisSer	280	
DB	838 ACCCCCAACATGATGAGCAACAAATCTTAAATCTGTTCAGAGAGAGCAAAAGTTTACACTCC	897	
QY	281 PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisProPheHis	300	
DB	898 TTTCAATAGCTTTTCTGTAGGGGAGATGTTTAAATATGACTGCTTCTCATCTTTTCAT	957	
QY	301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys	320	
DB	958 GCAAACCCCAACATTATATAGCGGAAGAACACAGAAACAGCTTATGACAAACAACCTTGT	1017	
QY	321 GlyProGlnCysTyrGlnHisIleLeuGlnGluGlyAlaLysGluPheAsnAlaAlaIleuThrAla	340	
DB	1018 GAAACAACAGTGTATCCAGCATTTTGAAGGAGCAAAAGAGTTTGTCTGTCTCTACCCGCT	1077	



QY 201 LysAspLeuGluAAspHisAspAspAspLeuSerArgProProArgLysPheProSer 220  
 DB 609 AAGAGTCTGGAGGATACCGAGATGATTAAGAAAGCGCCACCTCGGAAATTTCTCTTCT  
 QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
 DB 669 GATAAAATTTTGAAGCAATTTCTCAATGTTTCCAGATTAAGGGGACAGCAAGAAAGACTA  
 QY 241 LysGluLysGlyLysGluLeuThrGluGluGluGluLeuProGluAlaLeuProGluLys 260  
 DB 729 AAGGAAATATTAAGAACTACCGAAGAGAGCTCCAGGGGACCTTCTCTCGAATGT  
 QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerIleHisSer 280  
 DB 789 ACCCCCAACATAGATGAGCAAAATGCTAATCTGTTGAGAGAGCAAAAGCTTACATCC 848  
 QY 281 PheHISThrLeuPheCysArgArgCysPheLysGlyThrAspCysPheLeuHisIleProPheHis 300  
 DB 849 TTTCAATCGCTTTCTGTAGGCGAGATTTTAATGTGACTGCTTCTTCAATCCTTTTCAT  
 QY 301 AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 DB 909 GCACACCCCAACATTAATAGCGAAGAAACACAGAAACAGCTCTAGACAAACAACTTGT 968  
 QY 321 GlyProGlnCysTyrGlnHisIleGluGluGluValAlaLysGluPheAlaAlaAlaLeuThrAla 340  
 DB 969 GGACCAAGTGTTTACAGCATTTGGAGGAGCAAAAGAGTTTGCTGCTCTGACCGCT 1028  
 QY 341 GGUATGIIeLysThrProProLysArgProGluGluValGlnArgGluGluArgLeuProAsn 360  
 DB 1029 GAGCGGATTAAGAACCCCAACAAAGTCAGAGAGGCCCAAGAGAGAGCGCTTCCCAT  
 QY 361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380  
 DB 1089 AACGTTAGCAGGCCCAACGACCCCACTTAATGTGCTGGATCAAGATACAGACAGCT 1148  
 QY 381 AspArgGluAlaGlyThrGluThrGluGluGluAsnAsnAspLysGluGluGluLys 400  
 DB 1149 GATGGGAGAGCAGGAGCTGAACCGGGGAGAGAACATGATTAAGAAAGAAAGAGAGAG 1208  
 QY 401 LysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGlnThrProIleLysMet 420  
 DB 1209 AAAAGATGAACCTTCAGCTCTCTGAGCAAAATTCCTGGTCAAAACCAATAAAGATG 1268  
 QY 421 LysProAsnIleGluProGluGluAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 DB 1269 AAGCCAAATATTGAACCTCTGAGAAATGAGATGAGTGGTCTGAAGCTCAATGTTT 1328  
 QY 441 ArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460  
 DB 1329 AAGAGCTCATTTGGCACTTAATGACAAATTTCTGTGCAATTTGAGTTAATGGAGCC 1388  
 QY 461 LysThrCysArgGlnValTyrGluPheArgValLysGluSerSerIleIleAlaProAla 480  
 DB 1389 AAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGATCTAAGCATAGCTCCACT 1448  
 QY 481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuThrAla 500  
 DB 1449 CCGGCTGAGATGTGATCTCTCCAGAGAAAGAAAGAAAGAAACCGGTGTGGGCT 1508  
 QY 501 AlaHisCysArgLysIleGluLeuLysLysAspGlySerSerAsnHisValTyrAsnTyr 520  
 DB 1509 GCACAGTCAAGAAAGATACAGCTGAAGAAAGAGCGGCTCTCTTAACATGTTTAAACAAT 1568  
 QY 521 GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 DB 1569 CAACCTGTGATCATTCACAGCGGAGCTTGTGACAGTGTGCTTGCCCTTGGTGTATACACA 1628  
 QY 541 AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 DB 1629 AATTTTGTGAAGATTTTGTCAATGTAGTTCAAGATGTCAAAACCGCTTCCGGAGTGC 1688

QY 561 ArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeuAlaValaArgGluCys 580  
 DB 1689 CGTGCAGAAAGCAGTCAACCAACAGAGGCCCGGCTACCTGCTCCAGAGTGT 1748  
 QY 581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisIleThrAspSerLysAsnValSer 600  
 DB 1749 GACCTTACCTCTCTTCTTACTTGTGAGCGGCTCAACATTTGGGACATTAATAATGTGTC 1808  
 QY 601 CysLysAsnCysSerIleGlnArgGlySerLysLysIleLeuLeuAlaProSerAsp 620  
 DB 1809 TCGAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCATGATGCTGGCACCTTGAC 1868  
 QY 621 ValAlaGlyThrGlyIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 DB 1869 GTGGCAGCTGGGGGATTTTATTAAGATCTGTGAGAAATTAATTCATCTCAGAA 1928  
 QY 641 TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGlyLysValTyrAspLys 660  
 DB 1929 TACTGTGAGAGATTAATTTCTCAAGATGAACTACAGAGAGGGAAAGTATGATAA 1988  
 QY 661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValaAspAlaThrArgLys 680  
 DB 1989 TACATGTAGCAGCTTCTGTTCAACTTGAAACATGATTTTGTGTGATGCAACCGCAAG 2048  
 QY 681 GlyAsnLysIleArgPheAlaAsnHisSerValaAsnProAsnCysTyrAlaLysVal 699  
 DB 2049 GGTAAACAAATTCGTTTTCGAATCATTCGGTAATCAACTGCTATGCAAAAGTT 2105

RESULT 13  
 US-10-909-035-95  
 ; Sequence 95, Application US/10909035  
 ; Publication No. US20050136493A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rubin, Mark A.  
 ; APPLICANT: Chinaiyan, Arul M.  
 ; APPLICANT: Laxman, Bharathi  
 ; APPLICANT: Streekumar, Arun  
 ; TITLE OF INVENTION: AMACR Cancer Markers  
 ; FILE REFERENCE: UM-09098  
 ; CURRENT APPLICATION NUMBER: US/10/909, 035  
 ; CURRENT FILING DATE: 2004-07-30  
 ; NUMBER OF SEQ ID NOS: 123  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 95  
 ; LENGTH: 2512  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-909-035-95

Alignment Scores:  
 Pred. No.: 0 Length: 2512  
 Score: 3781.00 Matches: 698  
 Percent Similarity: 99.9% Conservative: 0  
 Best Local Similarity: 99.9% Mismatches: 1  
 Query Match: 99.3% Indels: 1  
 DB: 10 Gaps: 0

US-10-773-302-2\_copy\_1\_699 (1-699) x US-10-909-035-95 (1-2512)  
 QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysThrAspArgValLys 20  
 DB 10 ATGGGCCAGATGGGAGAAATCTGAGAGGACCAAGTTGTTGGCGGACGCTGTAATA 69  
 QY 21 SerGluTyrMetCysArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
 DB 70 TCAGAGTACAGCGACAGTGAAGCACTCAAGAGGTTCAAGACAGCTGATGAAGTAAAGGT 128  
 QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTyr 60  
 DB 129 ATGTTTATGTTCAATTCGACGAAATTTTGGAAAGAACGGAAATCTTAACCAAGATGG 188  
 QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80



Db 189 AACACGAGGATACAGCTGTGCATCTGACTTGTGAGCTATTGCGCGGACT 248  
 QY ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100  
 Db 249 AGGAGTGTTCGGTGCACGATGATGATTTCCAAACCAAGTCATCCCATTAAGACT 308  
 QY LeuAlaValAlaSerValProIleMetTyrSerTrpSerProLeuGlnIleAspPhe 120  
 Db 309 CTGATGCGAGTCTTCACTACCAACATATGTAATTCCTGCTCCCTACAGCAGAAATTT 368  
 QY MetValGluAspGluThrValIleuHisAsnIleProTyrMetGlyAspGluValIleuAsp 140  
 Db 369 ATGGTGAAGATGAAACTGTTTACATATCTCTTATATGGAGATGAAAGTTTATGAT 428  
 QY GlnAspGlyThrPheIleGluGlnLeuIleIleuAsnTyrAspGlyValHisGlyAsp 160  
 Db 429 CAGGATGTGACTTTCATTAAGAACTAAATAAATAATGATGGAAAGATACACGGGAT 488  
 QY ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGln 180  
 Db 489 AGAGATGTGGGTTTATTAATGATGAATTTTGTGAGATTGTGATCCCTTGCTCA 548  
 QY TyrAsnAspAspAspAspAspAspGlyAspAspProGluGlnArgGluGln 200  
 Db 549 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608  
 QY LysAspLeuGlnAspHisAspAspAspGlyGlySerAspProProArgLysPheProSer 220  
 Db 609 AAAGATCTGGAGGATCAACGAGATGATTAAGAAAGCCGCCACCTCGAAATTTCTCTCT 668  
 QY AspLysIlePheGlnAlaIleSerSerMetPheProAspLysGlyThrAlaGluGlnLeu 240  
 Db 669 GATTAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGCAACAGCAAGAACTA 728  
 QY LysGluLysTyrLysGluLeuThrGlnGlnLeuProGlyAlaLeuProProGluCys 260  
 Db 729 AAGGAAATAATTAAGAACTCAACGAGAGCTCCAGCGCACTTCTCTGATGT 788  
 QY ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerIleuHisSer 280  
 Db 789 ACCCCCACTATGATGAGCAACATGCTAAATCTGTTCAAGAGAGAGCAAGGCTTACACTCC 848  
 QY PheHisThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHisAspPheHis 300  
 Db 849 TTTCATAGCTTTTCTGAGCGATGTTTAAATAGTCTGCTCTTACATCTTTTCA 908  
 QY AlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeuAspAsnLysProCys 320  
 Db 909 GGAACACCCCACTATTAAGCGGAAGAACACAGAAACAGCTCTAGACAAACAACTTGT 968  
 QY GlyProGlnCysTyrGlnHisLeuGlnGlyAlaLysGluPheAlaAlaLeuThrAla 340  
 Db 969 GGAACACAGTGTACAGCATTTGGAGGAGCAAGAGATTTGCTGCTCTCAACCGCT 1028  
 QY GlnArgGlyIleValThrProProLysArgProGlyValArgArgArgGlyValArgLeuProAsn 360  
 Db 1029 GAGCGATTAAGACCCCAACCAACGCTCCAGAGGCGCGAGAGAGAGAGCGCTTCCCAT 1088  
 QY AsnSerSerArgProSerThrProThrIleAsnValLeuGlnSerLysAspThrAspSer 380  
 Db 1089 AACATTAAGAGCCCAACGACCCCAACATTAATGCTGCTGATCAAGATACAGACAGT 1148  
 QY AspArgGluAlaGlyThrGluThrGlnLysGluGluAsnAspLysGluGlnGluLys 400  
 Db 1149 GATAGGGAAGCAAGCACTGAACCGGGGAGAGAAACAATGATTAAGAAAGAAAGAGAG 1208  
 QY LysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGluThrProIleLysMet 420  
 Db 1209 AAAGATGAAGACTTCGAGCTCTCTGAGGAAATTCCTGGTGTCAACCAATTAAGAGAG 1268  
 QY LysProAsnIleGluProProGluLysAsnValGluTyrSerGlyAlaGluAlaSerMetPhe 440  
 Db 1269 AAGCCAAATATTAAGAACTCTCTGAGATGTGAGATGTGAGTGTGAGGCTCAATGT 1328

QY 441 ArgValIleuIleGlyThrTyrTrpAspAspPheCysAlaIleAlaArgLeuIleGlyThr 460  
 Db 1329 AGAGTCTCATTTGGAGCACTTACTATGACAAATTTCTGTCCATTTGCTGATTAATGGAGCC 1388  
 QY LysThrCysArgGlnValTyrGluPheArgValLysGlySerSerIleIleAlaProAla 480  
 Db 1389 AAAACATGATGACAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1448  
 QY ProAlaGluAspValAspThrProProArgLysLysLysLysLysLysLysLysLysLysLys 500  
 Db 1449 CCGGCTGAGATGTGGAATCTCTTCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1508  
 QY AlaHisCysArgLysIleGluLeuLysLysAspGlySerSerSerAsnHisValTyrAsnTyr 520  
 Db 1509 GCACACTGAGAAAGATACAGCTGAAAGAAAGAGCGCTCTCTTACCAATGTTTACAACTAT 1568  
 QY GlnProCysAspHisProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540  
 Db 1569 CAACCTGTGATCATTCCACGAGGCTGTGACAGTTGTGCTTGTGTATACACAA 1628  
 QY AsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsnArgPheProGlyCys 560  
 Db 1629 AATTTGTGAAGATTTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1688  
 QY ArgCysLysValGlnCysAsnThrLysGlnCysProCysTyrTrpLeuAlaValArgGluCys 580  
 Db 1689 CGCTGCAAGCAACATGACCAACGAGGCTGTGACAGTTGTGCTTGTGTATACACAA 1748  
 QY AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600  
 Db 1749 GACCTGACCTCTGTCTTACTTGTGAGCCCGTGACATTTGGGACAGTAAATGTGTCTC 1808  
 QY CysLysAsnCysSerIleGlnArgGlySerLysLysLysLysLysLysLysLysLysLysLys 620  
 Db 1809 TGCAAGAACTGACAGTATCAAGCGGGGCTCAAAAGACATCTATTCCTGCGACCATCTGAC 1868  
 QY ValAlaGlyTyrPheIlePheIleLysAspProValGlnLysAsnGluPheIleSerGlu 640  
 Db 1869 GTGGAGGCTGGGGATTTTATCAAAAGATCTGTGAGAAATGAATTCATCTCAGAA 1928  
 QY TyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgLysValTyrAspLys 660  
 Db 1929 TACTGTGAGATTAATTTCTCAAGATGAAGCTGACAGAAAGGAAAGTGTATGATTA 1988  
 QY TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAspAlaThrArgLys 680  
 Db 1989 TACATGTCAAGCTTCTGTTCAACTGAACATGATTTTGTGTGATGCAACCGCAG 2048  
 QY GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCysTyrAlaLysVal 699  
 Db 2049 GGTAACAAAATTCGTTTGAATCAATTCGTTAAATCCAAATCTGTAAGCAAAAGTT 2105

## RESULT 14

US-10-104-047-1192  
 ; Sequence 1192, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: NO. US20030236392A1 full length cDNA  
 ; FILE REFERENCE: HI-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1192  
 ; LENGTH: 2476  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-1192

## Alignment Scores:

Pred. No.:	0	Length:	2476
Score:	3569.50	Matches:	659
Percent Similarity:	94.3%	Conservative:	0
Best Local Similarity:	94.3%	Mismatches:	1
Query Match:	93.7%	Indels:	39
DB:	7	Gaps:	1

US-10-773-302-2\_copy\_1\_699 (1-699) x US-10-104-047-1192 (1-2476)

```
QY      1 MetGlyGlnThrGlyLeuValSerGlnuValProValCysTrpArgValValLeu 20
DB      93 ATGGGCCAGACTGGAGAAATCTGAGAGGAGCCAGTTGTGGCGGAGACGGGTAAA 152
QY      21 SerGluTrpMetCArgLeuArgGlnLeuValSerPheArgValArgAlaAspGluValLeuSer 40
DB     153 TCAGAGTGCATGCCAGCTGAGACAGCTCAAGAGTTCCAGACAGCTGAGAGGAAATGCT 212
QY      41 MetPheSerSerAspAspArgGlnuValLeuGluValArgThrGluValLeuSerGlnGluTrp 60
DB     213 ATGTTTAGTCCAAATCGTCAGAAAATTTTGGAAAGAACGGAATCTTAAACCAAGAAATGG 272
QY      61 LysGlnuArgArgGlnLeuProValHisLeuThrSerValSerSerLeuArgGlyThr 80
DB     273 AAACGCGAAGATACAGCTGTGTGCACATCCGACTTCTGTAGCTCATCGCGGAGCT 332
QY      81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValLeuProLeuLysThr 100
DB     333 AGGAGG----- 338
QY     101 LeuAsnAlaValAlaSerValProLeuMetYrSerTrpSerProLeuGlnLeuAsnPhe 120
DB     338 ----- 338
QY     121 MetValGluAspGlnThrValLeuHisAsnGlnProTrpMetGlyAspGluValLeuAsp 140
DB     339 ---GGGAGAGAGAACTGTTTACATACATCTCTTAATGAGAGATGAAAGTTTAAAT 395
QY     141 GlnAspGlyThrPheLeuGlnuLeuValLeuValAsnTrpAspGlyLeuValHisGlyAsp 160
DB     396 CAGAGTGTACTTTTCATTGAAGAACTAATTAATAAATTTATGATGGAAAGTACACGGGAT 455
QY     161 ArgGluCysGlyPheLeuAsnAspGluValPheValGluLeuValAsnAlaLeuGln 180
DB     456 AGAGATGTGGGTTTAAATGATGAAATTTTGTGGAGTGTGTAATGCCCTTGCTCAA 515
QY     181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGlnuArgGluGluValGln 200
DB     516 TATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575
QY     201 LysAspLeuGlnuAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220
DB     576 AAAGATCTCGAGAGATCACCGAGATGATAAAGAAACCCGCCACCTCGGAAATTTCTTCT 635
QY     221 AspLysIlePheGlnuAlaIleSerSerMetPheProAspLysGlyThrAlaGlnuLeu 240
DB     636 GATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGACGAGAAACTCA 695
QY     241 LysGlnuValSerGlnuLeuThrGlnuGlnuLeuProGlnuValaLeuProProGluCys 260
DB     696 AAGGAAATATTAAGAACTCACCGAAGAGAGCTCCAGGCGCATTTCTCTCTGAAATGT 755
QY     261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnuArgGlnuGlnuSerLeuHisSer 280
DB     756 ACCCCCAACATGAGTGAAGCAAAATGCTAAATCTGTTGAGAGAGACAAAGCTTACACCTC 815
QY     281 PheHisThrLeuPheCysArgArgCysPheLysTrpAspCysPheLeuHisProPheHis 300
DB     816 TTTCAATAGCTTTTCTGTAGGCGATGTTTAAATGATGATGATGATGATGATGATGATGAT 875
QY     301 AlaThrProAsnThrTrpLysArgLysAsnThrGlnuThrAlaLeuAspAsnLysProCys 320
DB     876 GCAACACCAACACTTATTAAGCGGAGAAACAGAAACGCTTAGACAAACAACTTGT 935
```

```
QY     321 GlyProGlnuCysTrpGlnuHisLeuGlnuGlyAlaLysGlnuPheAlaAlaAlaLeuThrAla 340
DB     936 GGAACCAAGTGTAAACGACATTTGGAGGGAGCAAAAGAGATTTGTGTGCTGTCCACCGCT 995
QY     341 GlnuArgIleLysThrProProLysArgProGlnuGlyValArgArgArgGlyArgLeuProAsn 360
DB     996 GAGCGATTAAGAACCCCAACCAAAACGTTCCAGAGGCGCGAGAGAGAGAGAGAGAGAGAG 1055
QY     361 AsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspThrAspSer 380
DB     1056 AAGCATGAGAGGCGCCCAACACCCCAACATTATGCTGTGAATCAAGATACAGACAGT 1115
QY     381 AspArgGlnuAlaGlyThrGlnuThrGlyGlnuAsnAsnAspLysGlnuGlnuLys 400
DB     1116 GATAGGAGAGCGAGGACTGAACCGGGGGAGAGAAATGATTAAGAAAGAAAGAGAGAG 1175
QY     401 LysAspGlnuThrSerSerSerSerGlnuAlaAsnSerArgCysGlnuThrProIleLysMet 420
DB     1176 AAAGATAAACTTGACCTCTCTGAAGCAAAATTTCTGGGTCTAAACCAATTAAGATG 1235
QY     421 LysProAsnIleGluProProGlnuAsnValGluTrpSerGlyAlaGluAlaSerMetPhe 440
DB     1236 AAGCCAAATATGTAACCTCTGAGAAATGTGAGAGTGAAGTGTGAAGCTCAATGCTT 1295
QY     441 ArgValLeuIleGlyThrTrpTrpAspAsnPheCysAlaIleAlaArgLeuIleGlyThr 460
DB     1296 AGAGTCTCATGTGCATCTACTATGACAAATTTCTGTGCTCAATGCTTAAGTTTGGAGCC 1355
QY     461 LysThrCysArgGlnuValTrpGluPheArgValLysGluSerSerIleIleAlaProAla 480
DB     1356 AAACATGTAACAGAGGTATGAGTTAGAGTCAAAAGAACTTAGACATCAATAGCTCCAGCT 1415
QY     481 ProAlaGluAspValAspThrProProArgLysLysLysArgLysHisArgLeuTrpAla 500
DB     1416 CCGCTGAGGATGTGATGATCTCTCAAGGAAAGAAAGAGAAACACCGGTGTGGGCT 1475
QY     501 AlaHisCysAspGlyIleGlnuLeuLysLysAspGlySerSerAsnHisValTrpAsnTrp 520
DB     1476 GCACATGCGAAGAAAGATACAGCTGAAGAAAGAGCGGCTCTTAACCAAGTTTAACACTAT 1535
QY     521 GlnProCysAspHisAspProArgGlnProCysAspSerSerCysProCysValIleAlaGln 540
DB     1536 CAACCTCGTGAATCATCCAGCGACCTTTGAGACGTTCTGTGCTGTGTGATAGCACAA 1595
QY     541 AsnPheCysGlnuLysPheCysGlnuCysSerSerGluCysGlnuAsnArgPheProGlyCys 560
DB     1596 AATTTTGTGAAAAGTTTGTCAATGTATGATTCAGAGTCAAAACCGCTTCCGGAGATGC 1655
QY     561 ArgCysLysAlaGlnuCysAsnThrLysGlnuCysProCysTrpLeuAlaValArgGluCys 580
DB     1656 CGCTGCAAGGACAGTCAACCAACAGAGCGCGGTCAACCTGTGCTCCGAGAGGT 1715
QY     581 AspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAspSerLysAsnValSer 600
DB     1716 GACCTGACCTCTGTCTTACTGTGAGACGCTCAACATTTGGACAGTAAATATGTGTCC 1775
QY     601 CysLysAsnCysSerIleGlnuArgLysLysLysHisLeuLeuLeuAlaProSerAsp 620
DB     1776 TGCAAGAACTGCAATATTCAGCGGGGCTCCAAAAGCATCTAATGTGGCAGCATGTGAC 1835
QY     621 ValAlaGlyTrpGlyIlePheLeuLysAspProValGlnuLysAsnGlnuPheIleSerGln 640
DB     1836 GTGGCAGGCTGGGGGATTTTATCAAAAGATCTGTGCAAAATTAATCATCTTCGAA 1895
QY     641 TyrCysGlnuGluIleIleSerGlnuAspGlnuAlaAspArgArgGlyLysValTrpAspLys 660
DB     1896 TACGTGAGAGATTAATTTCTCAAGATGAAGCTCAACAGAGAGGAAAGTATGATTA 1955
QY     661 TyrMetCysSerPheLeuPheAsnLeuAsnAspPheValValaAspAlaThrArgLys 680
DB     1956 TACATGTGAGACTTCTGTCAACTTGAACAATGATTTTGTGTGTGATGCAACCGGCAAG 2015
```



```
|||||
Db 1356 AAACATGTACACAGGTATGATGTTAGATCAAAAGATCTAGCATCATAGCTCCAGCT 1415
Qy 481 ProAlaGluAapValAapThrProProArgLysLysLysArgLysHisArgLeuThrAla 500
Db 1416 CCCGTGAGATGTGGATATCTCTCCAAAGAAAAAGAGAGAAACACCGGTGTGGCT 1475
Qy 501 AlAHsCySAArgLysIleGlnLeuLysLysAapGlySerSerAsnHisValTyrAsnTyr 520
Db 1476 GCACACTGCAGAAAGATACAGCTGAAAAAGACGGCTCTTAACCAAGTTTACACAT 1535
Qy 521 GlnProCyAapHisAProArgGlnProCyAapSerSerCyAProCyValIleAlaGln 540
Db 1536 CAACCTGTGATCATCCACGCGACGCTGTGACAGTTGTCGCCCTTGTGTATGACCAA 1595
Qy 541 AsnPheCySGluLysPheCySGlnCySerSerGluCySGlnAsnArgPheProGluCyS 560
Db 1596 AATTTTGTGAAAAGTTTGTCAATGTAGTCAAGTGTCAAAACCGCTTCCGGGATGC 1655
Qy 561 ArgCySLysAlaGlnCyAsnThrLysGlnCySProCySTyrLeuAlaValArgGluCyS 580
Db 1656 CGCTGCAGAACGACAGTGCACACCAAGCAGTCCCGTGTACTGTGCTGCAGAGTGT 1715
Qy 581 AapProAapLeuCySLeuThrCySGlyAlaAlaAapHisTrpAapSerLysAsnValSer 600
Db 1716 GACCTGTACTGTGTACTGTGTGAGCCGCTGACCATGTGGACAGTAAAAATGTGTCC 1775
Qy 601 CysLysAsnCySerIleGlnArgGlySerLysLysHisLeuLeuAlaProSerAap 620
Db 1776 TGCAAGAACTGCAGATTCAGGGGGCTCCAAAAGACATCTATTGTGGCACCATCTGAC 1835
Qy 621 ValAlaGlyTTPGlyLysPheLysAapProValGlnLysAsnGluPheIleSerGlu 640
Db 1836 GTGGAGGCTGGGGATTTTATCAAGATCCTGTGCAGAAAAAATGAATTCATCTCAGAA 1895
Qy 641 TyrCySGlyGluIleIleSerGlnAapGluAlaAapArgArgGlyLysValTyrAspLys 660
Db 1896 TACTGTGAGAGATTAATTCACAGATGAGCTGACAGAGGGGAAAGTGATGATAAA 1955
Qy 661 TyrMetCySerPheLeuPheAsnLeuAsnAapPheValValAspAlaThrArgLys 680
Db 1956 TACATGTGCAGCTTCTGTCAACTTGAAACAATGATTTTGTGGTGAATGCAACCCGCAAG 2015
Qy 681 GlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCySTyrAlaLysVal 699
Db 2016 GGTAAACAATTCGTTTGCAAATCATTCGTAAATCCAACTGCTATGCAAAAGTT 2072
```

Search completed: August 13, 2006, 18:01:50  
Job time : 2084.88 secs



```
Db      1 ATGGCCAGACTGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGAACGTGTA 60
Qy      21 SerGluYrMetArgLeuArgGluLeuYsArPheArgAlaAspGluValIysSer 40
Db      61 TCAAGATACATGCGATGAGACGCTCAAGAGCTTCAACAGACTGATGATGAAGAT 120
Qy      41 MetPheSerSerAsnArgGlnIleuGluValArgThrGluIleuAsnGlnGluTrp 60
Db      121 ATGTTTGGTTCCATTCGTCAAGAAATTTTGGAAAGAACGAAATCTTAAACCAAGAA 180
Qy      61 LysGlnArgArgGlnIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80
Db      181 AAACGCGCAAGAGATACAGCCCTGTGCACATCTGTGAGCTCATTCGGGGAGCT 240
Qy      81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuIleThr 100
Db      241 AGGAGATGTTGGTGACCAAGTGAATTTCCACACAGATCATCCATTAAAGACT 300
Qy      101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnIleAsn 120
Db      301 CTGAATGCAAGTTGCTTCAAGTACCAATATGATTTCTTGCTCCCTTACAGACAAATTT 360
Qy      121 MetValGluAspGluThrValLeuHisAsnIleProGlyYrMetGlyAspGluValLeuAsp 140
Db      361 ATGGTGGAAAGATGAACCTGTTTACATTAACATTCCTTATATGGGAGATGAAGTTTACAT 420
Qy      141 GlnAspGlyThrPheIleGluIleuIleYsAsnYrAspGlyLysValHisGlyAsp 160
Db      421 CAGATGCTACTTTTCATTGAAAGAACTAATAAAATTTGATGGAAAGTACACGGGAT 480
Qy      161 ArgGluCysGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180
Db      481 AGAGATGTGGGTTTATTAATGATGAATTTTGTGGAGTGGTGAATGCCCTTGCTGAA 540
Qy      181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluIleGluIleGln 200
Db      541 TATATATGATGATGACGATGATGATGATGACGATCTCTGAAGAAAGAAAGAAAGAGAG 600
Qy      201 LysAspLeuGluIleAspHisArgAspAspLysGluSerArgProProArgLysPheProSer 220
Db      601 AAAGATCTGGAGATACCCGAGATATTAAGAAAGCCGCCCACTCGAATATTCCTTCT 660
Qy      221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluIleu 240
Db      661 GATAAATATTTTGAAGCCATTTCTCAATGTTTCCAGATTAAGGGCAGCAGAAAGACTA 720
Qy      241 LysGluIleYrLysGluLeuThrGluGlnIleuProGlyValAlaLeuProProGluCys 260
Db      721 AAGGAAAAATATMAAGAACTCACCAACAGCAGCTCCAGGGGCACTTCTCTGAATGT 780
Qy      261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerIleuHisSer 280
Db      781 ACCCCCAATATGATGACCAATCTTAATCTGTTCAGAGAGCAAAAGCTTACATCC 840
Qy      281 PheIleThrLeuPheCysArgArgCysPheLysYrAspCysPheLeuHis----- 297
Db      841 TTTTATACGCTTTTCTGATAGCGATGTTTAAATATGACTGCTTCTCATCATGTAAGTGC 900
Qy      298 -----ProPheHisAlaThrProAsnThrYrLysArgLysAsnThrGluThrAlaLeu 315
Db      901 AATTATTCCTTTTCATGCAACACCCAACTTATTAAGCGAGAAACACAGAAACACCTCTA 960
Qy      316 AspAsnLysProCysGlyProGlnCysYrGlnHisLeuGluGluValAlaLysGluPheAla 335
Db      961 GACACAAACCTTTGGAGACCAAGTGTTCACAGCATTTGGAGAGGAGCAAAAGAGATTGCT 1020
Qy      336 AlaAlaLeuThrAlaGluArgIleLysThrProProLysArgProGlyValArgArgArg 355
Db      1021 GCTGCTTCACCGCTGAGGAGGATTAAGACCCCAACAAACGTCAGAGGCGGAGAGAGA 1080
Qy      356 GlyValGluLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375
Db      1081 GGACGCGCTTCCCAATTAACATGAGAGGCGCCACACCCCAACCAATTAAATGTGCTGAATCA 1140

Qy      376 LysAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGluIleAsnAsnAspLys 395
Db      1141 AAGGATACAGACAGTGTATGGAAGCAGGAGACTGAACCGGGGAGAGAAACAATGATTA 1200
Qy      396 GluGluGluGluLysLysAspGluThrSerSerSerSerSerGluIleAsnSerArgCysGln 415
Db      1201 GAAGAAAGAAAGAAAGAAAGTGAACCTTCAAGCTCTCTAAGTGAAGAAATTCCTGGTCCA 1260
Qy      416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAla 435
Db      1261 ACACCAATTAAGATGAAGCCAAATATTTGAACCTCTGAGATGTGAGTGGAGTGGTGCCT 1320
Qy      436 GluAlaSerMetPheArgValLeuIleGlyThrYrYrAsnAsnPheCysAlaIleAla 455
Db      1321 GAAGCTCAATGTTTAAAGTCTTCATAGGCACTTACTATGACAATTTCTGTGCATTTGCT 1380
Qy      456 ArgLeuIleGlyThrLysThrCysArgGlnValYrGluPheArgValLysGluSerSer 475
Db      1381 AGGTTAATGGGACCAAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAAAGAAATTCAGC 1440
Qy      476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysValArgLys 495
Db      1441 ATCATACCTCAGCTCCCGCTGAGAGATGTGATCTCTCCAGAGAAAGAAAGAGAGAA 1500
Qy      496 HisArgLeuThrAlaAlaHisCysArgLysIleGlnLeuLysLysAspGlySerSerAsn 515
Db      1501 CACCGCTTGGGGGTGCACACTGCAGAAAGATACAGCTGAAGAAAGAGAGCGCTCTCTTAC 1560
Qy      516 HisValYrYrAsnYrGlnProCysAspHisProAspGlnProCysAspSerSerCysPro 535
Db      1561 CATTTTACCACTATCAACCTGTGATCATTCACAGGAGCCTGTGTGACAGTTCTGCTCCCT 1620
Qy      536 CysValIleAlaGlnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 555
Db      1621 TGTGTGATGACCAAAATTTTGTGAAAGATTTTGTCAATGTAGTTTCAAGTGTCAAAAC 1680
Qy      556 ArgPheProGluCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysYrLeu 575
Db      1681 CGCTTTCGGGATGCGCTGCAAGCAGAGTGAACCAAGCAGTCCCGCTGCTACCTG 1740
Qy      576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisLeuAsp 595
Db      1741 GCTGTCCGAGAGTGAACCTGACCTGTCTTACTGTGTGAGCGGCTGACCATTTGGGAC 1800
Qy      596 SerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySerLysLysLeuLeu 615
Db      1801 AGTAAAAATGTGTCTGCAAGAACTGCAGATTTCAAGCGGGGCTCCAAAAGCATCTATTG 1860
Qy      616 LeuAlaProSerAspValAlaGlyYrTrpGlyIlePheIleLysAspProValGlnLysAsn 635
Db      1861 CTGGCACCATGTGACGTGGCGAGCTGGGGGATTTTATTAAGATTCCTGTGCAAAAAAT 1920
Qy      636 GluPheIleSerGluYrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGly 655
Db      1921 GAATTCATCTCAGAAATCTGTGAGAGATTAATTTCTCAAGTGAAGCTGACAGAGAGGG 1980
Qy      656 LysValYrAspLysYrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValVal 675
Db      1981 AAAGTGTATGATTAATACATGTGACAGCTTTGTGTCAACTTGAACATGATTTGTGGTG 2040
Qy      676 AspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695
Db      2041 GATCAACACCCGCAAGGATTAACAAATTCGTTTGCATATCATTCGGTAAATCCAAATCGC 2100
Qy      696 TyrAlaLysVal 699
Db      2101 TATGCAAAAGTT 2112

RESULT 2
US-11-266-748A-25769
; Sequence 25769, Application US/11266748A
; Publication No. US20060134663A1
```

```

: GENERAL INFORMATION:
: APPLICANT: Hartkin, Paul
: APPLICANT: Johnston, Patrick
: APPLICANT: Mulligan, Karl
: TITLE OF INVENTION: Transcriptional Microarray Technology and
: TITLE OF INVENTION: Methods of Using the Same
: FILE REFERENCE: 55815-0102 (319189)
: CURRENT APPLICATION NUMBER: US/11/566,748A
: PRIOR FILING DATE: 2005-11-03
: PRIOR APPLICATION NUMBER: EP 04105479.2
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105482.6
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105483.4
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105507.0
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105485.9
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: EP 04105484.2
: PRIOR FILING DATE: 2004-11-03
: PRIOR APPLICATION NUMBER: US 60/662,276
: PRIOR FILING DATE: 2005-03-14
: PRIOR APPLICATION NUMBER: US 60/700,293
: PRIOR FILING DATE: 2005-07-18
: NUMBER OF SEQ ID NOS: 483996
: SOFTWARE: PatentIn version 3.3
: SEQ ID NO 25769
: LENGTH: 2655
: TYPE: DNA
: ORGANISM: Homo Sapiens
US-11-266-748A-25769

```



QY 516 HisValIYrAsnYrGlnProCYsAspHisProArgGlnProCYsAspSerSerCyPro 535  
| | | | |  
Db 1683 CATGTTTACACTATCAACCTGTGATCATCCAGGACCTTGTGACATTTGTCGCCCT 1742  
QY 536 CyValIleAlaGlnAsnPhcYsGluLysPhcYsGlnCySerSerGlnCYsGlnAsn 555  
| | | | |  
Db 1743 TGTGTGATACACAAATTTTGTGAAAGTTTGTCAATGTAGTTGACAGTGCANAAC 1802  
QY 556 ArgPheProGlyCYsAspCYsLysAlaGlnCYsAsnThrLysGlnCYsProCYsYrLys 575  
| | | | |  
Db 1803 CGCTTTCCGGAGTGCCTGCGAAGACACAGTCAACACCAAGCATGCCGTGCTACCTG 1862  
QY 576 AlaValArgGluCYsAspProAspLeuCYsLeuThrCYsGluAlaAlaAspHisTrpAsp 595  
| | | | |  
Db 1863 GCTGTCCGAGAGTGTGACCTGACCTTGTCTTACTGTGAGGCCGTGACCAATTGGAC 1922  
QY 596 SerLysAsnValSerCYsLysAsnCYsSerIleGlnArgGlySerLysLysLysLeu 615  
| | | | |  
Db 1923 AGTAAATAATGTCTCTGCAAGAACCTGCAGTATTCAGCGGGGCTCAAAAAGCATCTATTG 1982  
QY 616 LeuAlaProSerAspValAlaGlyTrpGlyLysPheIleLysAspProValGlnLysAsn 635  
| | | | |  
Db 1983 CTGGCACCATCTGACGTCGGAGGCTGGGGATTTTATCAAGATCCCTGTGCAGAAAAT 2042  
QY 636 GluPheIleSerGluYrCYsGlyGluIleIleSerGlnAspGluAlaAspArgArgGly 655  
| | | | |  
Db 2043 GAATTCATCTCAGAAATCTGTGGAGATTTTCTCAAGATGAAGTCAGACAGAAAGGG 2102  
QY 656 LysValIYrAspLysYrMetCYsSerPheLeuPheAsnLeuAsnAspPheValI 675  
| | | | |  
Db 2103 AAAGTGTATGATTAATACATGTGACAGCTTCTGTCAACTGAACAATGATTTGTGTG 2162  
QY 676 AspAlaThrAspGlyGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCy 695  
| | | | |  
Db 2163 GATCAACCCCGCAAGGATGAACAAATTCGTTTGCAATCATTCGTAAATCCAAATGCC 2222  
QY 696 TyrAlaLysVal 699  
| | | | |  
Db 2223 TATGCAAAAGTT 2234

RESULT 3  
US-10-539-630-3  
; Sequence 3, Application US/10539630  
; Publication No. US20060104981A1

GENERAL INFORMATION:

; APPLICANT: Takeda Chemical Industries, Ltd.

; TITLE OF INVENTION: Preventing and treating agent for cancer

; FILE REFERENCE: 3130WOP

; CURRENT APPLICATION NUMBER: US/10/539,630

; PRIOR FILING DATE: 2005-06-17

; PRIOR APPLICATION NUMBER: JP2002-373144

; PRIOR FILING DATE: 2002-12-24

; NUMBER OF SEQ ID NOS: 14

; SEQ ID NO 3

; LENGTH: 2695

; TYPE: DNA

; ORGANISM: Human

US-10-539-630-3

Alignment Scores:

Pred. No.: 1,83e-310 Length: 2695

Score: 3787.50 Matches: 698

Percent Similarity: 99.1% Conservative: 0

Best Local Similarity: 99.1% Mismatches: 1

Query Match: 99.5% Indels: 5

DB: 6 Gaps: 1

US-10-773-302-2\_copy\_1\_699 (1-699) x US-10-539-630-3 (1-2695)

QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCYsTrpArgLysArgValLys 20  
| | | | |  
Db 167 ATGGCCAGACTGGGAGAAATCTGAGAGGACCAAGTTTGTGGCGGAAAGGTGTAAAA 226

QY 21 SerGluYrMetArgLeuArgGlnLeuLysArgPheArgAlaAspGluValLysSer 40  
| | | | |  
Db 227 TCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTGACAGAGCTGATGAAGTAAAGAGT 286  
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTrp 60  
| | | | |  
Db 287 ATGTTAATTCATGATGTCAGAAATTTTGGAAAGAACGGAATCTTAAACAAAGATGG 346  
QY 61 LysGlnArgArgIleGlnProValHisIleLeuThrSerValSerSerLeuArgGlyThr 80  
| | | | |  
Db 347 AAACACGAGAGATACAGCTGTCGATCTGACATCTTGTGTGAGTCACTTGGCCGGGACT 406  
QY 81 ArgGluCYsSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100  
| | | | |  
Db 407 AGGAGTGTCCGCTGACAGTGCATTTGATTTCCAAACAAAGTATCCCATTTAAAGACT 466  
QY 101 LeuAsnAlaValAlaSerValProIleMetYrSerTrpSerProLeuGlnGlnAsnPh 120  
| | | | |  
Db 467 CTGAATGCACTTCTTCAAGTACCACTAATGATTTGTGCTCCCTACAGCAAGATTTT 526  
QY 121 MetValGluAspGluThrValLysHisAsnIleProTYrMetGlyAspGluValLysAsp 140  
| | | | |  
Db 527 ATGTGGAAGATGAATGACTGTTTACATTAATTCCTTATATGGAGATGAAGTTTATGAT 586  
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTYrAspGlyLysValHisGlyAsp 160  
| | | | |  
Db 587 CAGGATGTACTTTCATTTGAAGAACTAATTAATAATATGATGGAAAGTACCGGGGAT 646  
QY 161 ArgGluCYsGlyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
| | | | |  
Db 647 AGGAATGTGGGTTTAAATGAATGAATTTTGGAGTTGGTGAATGCCCTTGTGCTCA 706  
QY 181 TyrAsnAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluLysGln 200  
| | | | |  
Db 707 TATATATATGATGACATGATGATGATGAGAGATCTGTAAGAAAGAAAGACAG 766  
QY 201 LysAspLeuGluAspHisArgAspAspLysGluSerArgProProAspGlyPheProSer 220  
| | | | |  
Db 767 AAAGATCTGAGGATGACCCGAGATGATTAAGAAAGCCGACCTCGAAATTTCTTCT 826  
QY 221 AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluLeu 240  
| | | | |  
Db 827 GATTAATTTTGAAGCATTTCTCAATGTTTCAATTAAGGACACAGCAAGAACTA 886  
QY 241 LysGluLysTYrLysGluLeuThrGluGlnGlnLeuProGlyAlaLeuProProGlyCy 260  
| | | | |  
Db 887 AAGGAAAATTAATAAGAACTCACCGAAGAGCTCCAGGCGCACTTCTCGAATGT 946  
QY 261 ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLeuHisSer 280  
| | | | |  
Db 947 ACCCCAAACATAGTAGACCAATGCTTAATCTGTTGAGAGAGCAAGGATTAACCTCC 1006  
QY 281 PheHisThrLeuPheCYsArgArgCYsPheLysTYrAspCYsPheLeuHis----- 297  
| | | | |  
Db 1007 TTTCATAGCTTCTTCTGAGCGAATGTTTAATATGACTGCTTCTTCAATCGTAAGTGC 1066  
QY 298 -----ProPheHisAlaThrProAsnThrTYrLysArgLysAsnThrGluThrAlaLeu 315  
| | | | |  
Db 1067 AATTATCTTTTATGACACACCAACCTTATTAAGCGAAGAACACAGAAACAGCTCTA 1126  
QY 316 AspAsnLysProCYsGlyProGlnCYsTYrGlnHisIleLeuGluGlyAlaLysGluPheAla 335  
| | | | |  
Db 1127 GACAAACAAACCTTGTGACCACTGTGTTACAGATTTTGGAGGAGCAAGAAAGATTTGCT 1186  
QY 336 AlaAlaLeuThrAlaGluArgIleLysThrProProLysArgProGlyGlyArgArg 355  
| | | | |  
Db 1187 GCTGCTCTCACCGCTGAGCGGATTAAGACCCCAACAAACGTTCCAGAGGCCGAGAAAG 1246  
QY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375  
| | | | |  
Db 1247 GAGCGCTTCCCATTAACATGATGACAGGCCACGCCACCATTAATGTGCTGGAATCA 1306

```

376 LysaSPThrAspSerAspArgGluAlaGlyThrGluThrGluGlyGluAsnAspLys 395
1307 AAGATACAGACAGATGATGAGAGACAGGACTGAAACGGGGAGAGACATATATAA 1366
396 GluGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415
1367 GAAAGAGAGAGAGAGAGATGAATCTTCAGACTCTCTCGAAGCAAAATCTCGGTCCA 1426
416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluThrSerGlyAla 435
1427 ACACCAATAAAGATAGAGCAAAATATGAACTCTCGAAGATGAGAGGTGTGTCT 1486
436 GluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla 455
1487 GAAGCTCAATGTTAAGTCTCTCATGCTGACCTTACATGACAAATTTCTGTCCATGCT 1546
456 ArgLeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer 475
1547 AGGTTAATTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAAATCTAGC 1606
476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysLysArgLys 495
1607 ATCATAGCTCCAGCTCCCGCTGAGATGTGATCTCCCTCCAAAGGAAAAAGAGGAAA 1666
496 HisArgLeuThrAlaAlaHisCysArgLysIleGluLeuLysLysAspGlySerSerAsn 515
1667 CACCGGTGTGTGGCTGACACATGCAAGAAAGATACAGCTGAAAAAGACGCGCTCTTAC 1726
516 HisValTyrAsnTyrGlnProCysAspHisProArgGlnProCysAspSerSerCysPro 535
1727 CATGTTTCAACTATCAACCCGTGTATCATCCAGCGAGCTTGTGACAGTTCGAGCTT 1786
536 CysValIleAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsn 555
1787 TGTGTGATAGACAAATTTTGTGAAAAGTTTGTCAATGATGATTCAGAGTGTCAAAAC 1846
556 ArgPheProGlyCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeu 575
1847 CGCTTCGGGAGTGGCGCTGCAAGACACAGTCAACACCAAGACAGTCCCGTCACTCG 1906
576 AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAlaAspHisTrpAsp 595
1907 GCTGTCAGAGAGTGTGACCCCTGACTCTGTCTTATCTTGTGAGACCGCTGACCATGGGAC 1966
596 SerLysAsnValSerCysLysAsnCysSerIleGlnArgGlySerLysLysIleLeuLeu 615
1967 AGTAAAAATGTGTCTGCAAGAACAGTATTCAGCGGGGCTCCAAAGATCATCTATTTG 2026
616 LeuAlaProSerAspValAlaGlyTyrGlyIlePheIleLysAspProValGlnLysAsn 635
2027 CTGGACCATCTGACGTGGCAGGCTGGGGGATTTTATCMAAGATCTGTGCAAGAAAT 2086
636 GluPheIleSerGluTyrCysGlyGlyIleIleSerGlnAspGluAlaAspArgArgGly 655
2087 GAATTCATCTCAGATATCTGTGAGAGATTAATTTCTCAAGATGAAGCTGACAGAGAGGG 2146
656 LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAsnAspPheValAla 675
2147 AAAGGTATGATTAATATACATGTGACAGCTTTCGTTCACTTGAAACAATGATTTGTGGTG 2206
676 AspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695
2207 GATGACACCCGCAAGGGTAAACAAATTCGTTTGTGAAATCATTCGGTAAATTCAAATGCG 2266
696 TyrAlaLysVal 699
2267 TATGCAAAAGTT 2278

```

RESULT 4  
 US-11-266-748A-29501  
 ; Sequence 29501, Application US/11266748A  
 ; Publication No. US2006013463A1  
 ; GENERAL INFORMATION:

```

APPLICANT: Harkin, Paul
APPLICANT: Johnston, Patrick
APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcription Microarray Technology and
TITLE OF INVENTION: Methods of Using the Same
FILE REFERENCE: 55815-0102 (319189)
CURRENT APPLICATION NUMBER: US/11/266,748A
CURRENT FILING DATE: 2005-11-03
PRIORITY APPLICATION NUMBER: EP 04105479.2
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: EP 04105482.6
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: EP 04105483.4
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: EP 04105507.0
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: EP 04105485.9
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: EP 04105484.2
PRIORITY FILING DATE: 2004-11-03
PRIORITY APPLICATION NUMBER: US 60/662,276
PRIORITY FILING DATE: 2005-03-14
PRIORITY APPLICATION NUMBER: US 60/700,293
PRIORITY FILING DATE: 2005-07-18
NUMBER OF SEQ ID NOS: 483996
SOFTWARE: PatentIn version 3.3
SEQ ID NO 29501
LENGTH: 2695
TYPE: DNA
ORGANISM: Homo Sapiens
US-11-266-748A-29501

Alignment Scores:
Pred. No.: 1,836-310 Length: 2695
Score: 3787.50 Matches: 698
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 99.5% Indels: 5
DB: 8 Gaps: 1

US-10-773-302-2_COPY_1_699 (1-699) x US-11-266-748A-29501 (1-2695)
QY 1 MetGlyGlnThrGlyLysLysSerGluLysGlyProValCysTrpArgLysArgValLys 20
DB 167 ATGGGCCAGCTGGAGAAATCTGAGAAAGGACCAAGTTGTGGGAGCGGTATAA 226
QY 21 SerGluTyrMetCysArgLeuArgGlnLeuLysArgPheArgValAlaAspGluValLysSer 40
DB 227 TCAGAGTACAGCCAGCTAGACAGCTCAAGAGTTCAACAGCTGATGAAGTAAAGGT 286
QY 41 MetPheSerSerAsnArgGlnLysIleLeuGluArgThrGluIleLeuAsnGlnGluTyr 60
DB 287 ATGTTAGTTCCAATCCGCAAGAAATTTTGAAAGAACGAAATCTTAAACCAAGAAATGG 346
QY 61 LysGlnArgArgGlnIleGlnProValHisIleLeuThrSerValLysSerLeuArgGlyThr 80
DB 347 AAACGCAAGAGATACAGCTGTGACACTTGTGAGTCAATTCGGCGGAGCT 406
QY 81 ArgGluCysSerValThrSerAspLeuAspPheProThrGlnValIleProLeuLysThr 100
DB 407 AGGAGAGTTCGGTGAACAGTGAAGCTGATTTTCCAAACAAGTATCCATTAAAGACT 466
QY 101 LeuAsnAlaValAlaSerValProIleMetTyrSerTyrSerProLeuGlnAsnPhe 120
DB 467 CTGAATGAGATGCTCAGTACCAATATGATTTCTGGTCTCCCTCAGCAGAAATTTT 526
QY 121 MetValGluAspGluThrValLeuHisAsnIleProTyrMetGlyAspGluValLeuAsp 140
DB 527 ATGTGGAGAGATGAACGTTTTCATTAACAATCTCTTATATGAGAGATGAAGTTTAAAT 586
QY 141 GlnAspGlyThrPheIleGluGluLeuIleLysAsnTyrAspGlyLysValHisGlyAsp 160
DB 587 CAGATGTATCTTTCATTTGAAGAACTAATAAATATATGATGGGAAAGTACACGGGAT 646

```

QY ArgGlucyGsglyPheIleAsnAspGluIlePheValGluLeuValAsnAlaLeuGlyGln 180  
DB AGAGATGTGGGTTTAAATGATGAAATTTTCTGAGTTGGGATGGCTTGGTCA 706  
QY TyrAsnAspAspAspAspAspAspGlyAspAspProGluGluGluGlyGln 200  
DB TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 766  
QY LysAspLeuGluAspPheIleAspAspAspGlySerProProAlaGlyPheProSer 220  
DB AAAGATCTGGAGGATCCGAGATGATTAAGAAAGCCGCACTCGGAAATTTCTCT 826  
QY AspLysIlePheGluAlaIleSerSerMetPheProAspLysGlyThrAlaGluGluLeu 240  
DB GATTAATTTTGAAGCATTTCTCTCATGTTTCCAGTAAAGGCAACAGCAAGAACTA 886  
QY LysGluLysTyrGlyGluLeuThrGluGlnGluLeuProGlyAlaLeuProGlyCys 260  
DB AAGGAAATATTAAGAATCAACCAAGAGAGCTCCAGGCGGCACTTCTCTGAATGT 946  
QY ThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerIleSer 280  
DB ACCCCCAACATGATGAGCAAAATGCTAAATCTGTTCAGAGAGCAAAAGCTTACACTCC 1006  
QY PheIleThrLeuPheCysArgArgCysPheLysTyrAspCysPheLeuHis----- 297  
DB TTTTCAATGCTTTTCTGTAGGCGATGTTTAAATTAATGATCTGCTTCTTCAATCGTAAGTCC 1066  
QY -----ProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeu 315  
DB AATTATCTTTTCAATGCAACCAACCACTTAAGCGGAAGAACACAGAAACAGCTCTA 1126  
QY AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluGluAlaLysGluPheAla 335  
DB GACACCAAACTTGTGAGCAACAGGTGTAACGCAATTTGAGGAGGCAAAAGATTTGCT 1186  
QY AlaAlaLeuThrAlaGluArgIleLysThrProProLysArgProGlyLysArgArgArg 355  
DB GCTGCTCTCAACCGCTGAGCGGATTAAGACCCCAACAAACGTCAGAGGCGGCAAGAA 1246  
QY GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375  
DB GACGCGCTTCCCAATTAACAGTAGAGAGCGCCAGACCCCAACATTAATGTCTGCAATCA 1306  
QY LysAspThrAspSerAspAspArgLysAlaGlyThrGluThrGlyGlyGluAsnAsnAspLys 395  
DB AAGGATACAGACAGTATGAGGAGAGCAAGCACTGAACGGGGGAGAGAACATGATATA 1366  
QY GluGluGluGluLysLysAspGluThrSerSerSerSerSerGluAlaAsnSerArgCysGln 415  
DB GAAAG 1426  
QY ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTyrSerGlyAla 435  
DB ACACCAATTAAGATGATGACCAAAATATTAACCTCTGAGAAATGTGAGATGAGTGTCT 1486  
QY GluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla 455  
DB GAAGCTCAATGTTTAAAGTCTCTCATGTGCACTTAAGCAATTTCTGTGCAATGCT 1546  
QY ArgLeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer 475  
DB AGGTTAATTTGGAGCCAAACATGTAACAGGTGATGATTTAGAGTCAAAAGATCTAGC 1606  
QY IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysLysArgLys 495  
DB ATCATAGCTCCAGCTCCGCTGAGATGTGATATCTCTCCCAAGCAAAAGAGAGAGAA 1666  
QY HisArgLeuThrAlaAlaHisCysArgArgLysIleGluLeuLysLysAspGlySerSerAsn 515  
DB CACCGGTGTGTGGCTGCACTGCAAGAAAGATACAGTGAAGAAAGAGAGAGAGAGAGAG 1726

QY HisValTyrAsnTyrGlnProCysAspPheIleProArgGluProCysAspSerSerCysPro 535  
DB CATGTTTAAACAATCATCAACCTGTGATCATCCAGGAGCTTGTGATGATGTGCTT 1786  
QY CysValIleAlaGlnAsnPheCysGluLysPheCysGlnCysSerSerGluCysGlnAsn 555  
DB TGTGTGTAGACCAAAATTTTGTGAAAGTTTGTTCATGTATGATTCAGAGTGTCAAAAC 1846  
QY ArgPheProGlyCysArgCysLysAlaGlnCysAsnThrLysGlnCysProCysTyrLeu 575  
DB CGCTTCCGAGAGCGCTCTCAAGCAAGTGCACCAAGAGAGAGAGAGAGAGAGAGAGAG 1906  
QY AlaValArgGluCysAspProAspLeuCysLeuThrCysGlyValAlaAspHisTyrAsp 595  
DB GCTGTCCGAGAGGTGACCTGTGACTTGTCTTAATTTGTGAGAGAGAGAGAGAGAGAG 1966  
QY SerLysAsnValSerCysLysAsnCysSerIleGlnArgLysSerLysHisLeuLeu 615  
DB AGTAAATATGTGCTTCCAG 2026  
QY LeuAlaProSerAspValAlaGlyTyrGlyTyrPheIleLysAspProValGluLysAsn 635  
DB CTGGCAACATCTGACGTCGAGGCTGGGAGATTTTATCAAGATCTGTGCAAGAAAT 2086  
QY GluPheIleSerGlyTyrCysGlyGluIleIleSerGlnAspGluAlaAspArgArgGly 655  
DB GAATTCATCTCAGAAATCTGTGAGAGATTAATTTCTCAAGATGAGCTGACAGAGAGAG 2146  
QY LysValTyrAspLysTyrMetCysSerPheLeuPheAsnLeuAsnAspPheValAla 675  
DB AAGGTATATGATTAATCATGTGCAAGCTTCTGTTCACCTGGAACATGATTTTGTGCTG 2206  
QY AspAlaThrArgLysGlyAsnLysIleArgPheAlaAsnHisSerValAsnProAsnCys 695  
DB GATCAACCCGCAAGGGTAAACAATAATCGTTTGCAATCATTCGTTAAATCCAAATCTCC 2266  
QY TyrAlaLysVal 699  
DB TATGCAAAAGTT 2278

RESULT 5  
US-11-266-748A-259286  
Sequence 259286, Application US/11266748A  
Publication No. US20060134653A1  
GENERAL INFORMATION:  
APPLICANT: Harkins, Paul  
APPLICANT: Harkins, Patrick  
TITLE OF INVENTION: Transcription Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 259286  
LENGTH: 1803

```

TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (356)..(379)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (1680)..(1680)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (1764)..(1764)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (1792)..(1792)
OTHER INFORMATION: n is a, c, g, or t
US-11-266-748A-259286

Alignment Scores:
Pred. No.: 1.78e-232 Length: 1803
Score: 2863.00 Matches: 549
Percent Similarity: 94.8% Conservative: 1
Best Local Similarity: 94.7% Mismatches: 18
Query Match: 75.2% Indels: 12
DB: 8 Gaps: 1

US-10-773-302-2_COPY_1_699 (1-699) x US-11-266-748A-259286 (1-1803)
QY 91 RHEPOTHRG1VAL11EPROLEUYSRTHLEUASPA1VAL1ASERVALPRO11MET 110
DB 82 TTCGCCAGCAAGTCATCCATTAAGACTCTGAATGCGAGTGTCTGAGTACCATAATG 141
QY 111 TYSERTRSEPRoleuGlnGlnAenPheMetValGluAspGluThrVal1LeuHisAen 130
DB 142 TATTCGTGCTCCCTTACGACGAAATTTATGCTGGAAGTAACTGTTTACCTAAC 201
QY 131 I1EPROTYMETG1YASPG1UVAL1LEUASPG1N1ASPG1YTHRPH11EG1UGLU1E 150
DB 202 ATTCCTTATATGAGGATGATGAAGTTTATAGTACAGATGTAATCTTCAATGAAGATTA 261
QY 151 LVASERTYRASPGLYLVSV1AH1EG1YASPAARGGLUCYAGGLYPHE11EASNPGLU1E 170
DB 262 AAAAATATGATGAGGAAAGTACACGGGATAGAGATGCGGTTTAAATGATGAAT 321
QY 171 PHEVALG1U1EUV1A1ASNA1ALEUG1YGIN1YR1ASNA1ASPA1ASPA1ASPA1 190
DB 322 TTGTGAGATTGCTGAATGCCCTTGTCTCAATATANNNNNNNNNNNNNNNNNNNGA 381
QY 191 AASPAPROGLUG1U1ARG1UG1U1YSG1U1YASPLEUG1U1ASPH11A1RG1ASPA1 210
DB 382 GACGATCCTGAAGAAAGAGAAAGAAAGAAAGATCTGAGAGATCAACGAGATGATAA 441
QY 211 GLUSERARGPROARG1YSPHEP1ROSERASP1Y11EPHEGLU1A11ESER1MET 230
DB 442 GAAAGCGCGCCACTCTCGAAATTTCTTCTGATTAATTTTGAACCAATTCCTCAAG 501
QY 231 PHEPROASPLV1EGL1YTH1A1AG1UG1U1EUV1YSG1U1YR1Y1YSG1U1EUTR1G1N 250
DB 502 TTTCAGATTAAGGCGCACAGCAAGAAAGCTTAAGGAAATTAAGAACTCACCGCAACG 561
QY 251 G1N1EUPROGLY1A1ALEUPROPROGLUCYR1THRPROASN11EASPG1YPROASN1 270
DB 562 CAGCTCCAGGCGCACTTCTCTCTGATATPACCCCAACATAGATGACCAATGCTAAA 621
QY 271 SERVALG1N1ARG1UG1N1SER1EUS1SER1PHE11RTH1LEUPHECY1ARG1YSPHE 290
DB 622 TCTGTTCAAGAGAGAGAAAGCTTACCTCTTTCATACGCTTTTCTGTAGGCGATGT 681
QY 291 LVETYRASP1CYSPHE1EUS1SPROPH11A1A1TH1PROASN1TH1YR1Y1Y1Y1Y1Y 310
DB 682 AATATATGACTGCTTCTTCAATCTTTTCATGCAACCCCAACCTTATAGCGGAAAGAAC 741

```

```

QY 311 THR1UTR1A1ALEUASPA1N1YSPROCYSG1YPROG1N1CYR1G1N1H1LEUG1UG1Y 330
DB 742 ACAGAAACAGCTCTTGAACAAACCTTGTGACCAAGTGTACAGATTTGAGAGGA 801
QY 331 A1ALYSG1UPHEA1A1A1ALEUTH1A1AG1UARG1Y1E1YR1THR1PRO1Y1Y1Y1Y1Y 350
DB 802 GCAAGAGATTTGCTGCTCTCTCAACCGCTGAGCGGATTAAGACCCCAACAAAGTCCA 861
QY 351 G1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 370
DB 862 GAGAGCGCCAGAGAGAGAGCGGCTTCCAAATTAACATAGACGCCACAGACCTT 921
QY 371 AENVAL1EUG1USER1YASPTHR1ASPSER1YASPA1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 390
DB 922 AATGTCTGAAATCAAGAGATTAACAGCTGATTAAGAGACGAGACTGAACCGGGGA 981
QY 391 G1U1A1EN1ASPA1YSG1UG1UG1UG1U1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 410
DB 982 GAGAACATGATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041
QY 411 AENSERARGYSG1N1THR1PRO1LEUYSRTHLEUASPA1VAL1ASERVALPRO11MET 430
DB 1042 AATTCGCGTGTCAAAACCAATTAAGATGAAGCCAAATATTGAACCTCTGAGAAATG 1101
QY 431 G1UTR1P1SERG1Y1A1AG1U1A1SER1ETR1PHEARG1Y1EGL1YTH1YR1Y1Y1Y1Y 450
DB 1102 GAGTGAAGTGTGTGAAGACCTCAAGTTTAAAGTCTCATTTGCGACTTACATAGCAAT 1161
QY 451 PHECYA1A1A1A1A1ARG1LEU11EGL1YTH1YR1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 470
DB 1162 TTCTGTGCAATTCGTAAGTATATGGAACCAACATGTAAGACAGGTGATAGATTAGA 1221
QY 471 VAL1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 490
DB 1222 GTCAAGAAATCTAGATCATAGCTCAGCTCCCGCTGAGAGTGTGATCTCTCCAAAG 1281
QY 491 LVSV1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 510
DB 1282 AAAAAGAAAGAGAAACCCGATGTGGGTGACCTGACAGAAAGATTAAGCTTGA 1341
QY 510 AASPGLYSER1SER1AEN1H1SVAL1YR1ASNTYR1G1N1YR1Y1Y1Y1Y1Y1Y1Y1Y 530
DB 1342 GAGCGGCTCTTAACATGTTTACATGATCAACCTGTATTCATCCACGGACGCTT 1401
QY 530 YASPSER1SER1CYSPROCYVAL11E1A1AG1N1ASNPHECY1GLU1YSPHECYSG1N 549
DB 1402 GTGACAGTTCCGCTTGTGTGATAGCAAAATTTTGTGAAAGATTTGTCAATGT 1461
QY 550 SER1SER1GLUCYSG1N1AEN1YR1PHEPROGLYCYARGCYVAL1Y1Y1Y1Y1Y1Y1Y 569
DB 1462 AGTTCAGAGTGTCAAAACCCGCTTCCCGAGATGCGCTGCAAGACAGTCAACACA 1521
QY 569 YSG1N1CYSPROCYR1YR1LEU1A1A1ARG1UCYASPPROASPLEUCY1EUTR1Y1Y 588
DB 1522 AGCATGCGCGTGTACCTGTGCTGTGAGAGTGTGACCTGACCTCTGTCTTACTGT 1581
QY 589 G1Y1A1A1A1ASPH11ETR1ASPSER1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 608
DB 1582 GGAGCGCGTCAATTTGGGACAGTAAATGTCTCTGCAAGAACTGCAAGTATTCAGCGG 1641
QY 609 G1Y1SER1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 628
DB 1642 GCGTCCAAAGAACATCTATTTGTGCAACCATCTGACG1N1G1Y1Y1Y1Y1Y1Y1Y 1701
QY 628 E1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 646
DB 1702 CAAGAGTCTGTGCAAGAAATTAATTTTCATCTTCAAGATATCTGTGGAGAGATTA 1761
QY 647 ----SERG1N1ASPG1U1A1ASPA1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y1Y 658
DB 1762 CTNCAAGTGAAGAGTGAACAGATGAGGGGAAAGTGTAT 1801

```



QY 510 SAAPGlySerSerAsnHisValTyrAsnTyrGlnProCysAsp-HisProArgGlnProC 530  
 DB 1342 GAACGGCTCTCTTAACCATGTTTACACATCAACCTTGATTTCAACGCGACCTT 1401  
 QY 530 YASPSerSerCysProCysValIleIleGlnAsnPheCys-GlnLysPheCysGlnCys 549  
 DB 1402 GTGACAGTTCGTGCCCTTGATGATGACAAATTTTGGTGAAGTGGTGTCAATGT 1461  
 QY 550 SerSer-GluCysGlnAsnArg-PheProGlyCysArgCysLysAlaGlnCysAsnThrL 569  
 DB 1462 AGTTCAGAGGTGTCAAACCCGCTTCCCGGATGCGCGTCAAGACAGTCACACCA 1521  
 QY 569 YAGL-CysProCysTyr-LeuAlaValArgLysAspProAspLeuLysLeuThrCys 588  
 DB 1522 ACCAGTCCCGGTGCTACCTGCGCTGTCAGAGGTGACCCGACCTCTGTCTTACTGT 1581  
 QY 589 GJYAlaAlaAspHisTyrAspSerLysAsnValSerCysLysAsnCysSerIleGlnArg 608  
 DB 1582 GGAGCGCGTGAACCATGGGACAGTAAATGTGTCTCGAAGAACTGCGATTTCAAGCG 1641  
 QY 609 GlySerLysLysHisLeuLeuLeuAlaProSerAspVal-AlaGlyTyrGlyTlePheI 628  
 DB 1642 GGCTCCAAAGACATCTATGCTGGCAGCATCTGACGTGCGAGCGTGGGATTTTAT 1701  
 QY 628 eLysAspProValGlnLysAsnGluPhe-IleSerGluTyrCys-GlyGlnIleIle-- 646  
 DB 1702 CAAAGATCTGTGCGAAGAAATGAAATTCATCTTCAGAACTGTGGGAGATTAATTTT 1761  
 QY 647 ---SerGlnAspGlnAlaAspArgArgGlyLysValTyr 658  
 DB 1762 CTNCAAGGTGAGAGGTGACACATGAGGAGAAAGTGTAT 1801  
 RESULT 7  
 US-11-266-748A-319803/c  
 ; Sequence 319803, Application US/11266748A  
 ; Publication No. US2006013463A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hatkin, Paul  
 ; APPLICANT: Johnston, Patrick  
 ; APPLICANT: Mulligan, Karl  
 ; TITLE OF INVENTION: Transcription Microarray Technology and  
 ; TITLE OF INVENTION: Methods of Using the Same  
 ; FILE REFERENCE: 55815-0102 (319189)  
 ; CURRENT APPLICATION NUMBER: US/11/266,748A  
 ; PRIOR APPLICATION NUMBER: EP 04105479.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105482.6  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105483.4  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105507.0  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105485.9  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105484.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: US 60/662,276  
 ; PRIOR FILING DATE: 2005-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/700,293  
 ; PRIOR FILING DATE: 2005-07-18  
 ; NUMBER OF SEQ ID NOS: 48396  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 319803  
 ; LENGTH: 1803  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (12)..(12)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:

NAME/KEY: misc\_feature  
 LOCATION: (40)..(40)  
 OTHER INFORMATION: n is a, c, g, or t  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (124)..(124)  
 OTHER INFORMATION: n is a, c, g, or t  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1425)..(1448)  
 OTHER INFORMATION: n is a, c, g, or t  
 US-11-266-748A-319803  
 Alignment Scores:  
 Pred. No.: 1.78e-232 Length: 1803  
 Score: 2863.00 Matches: 549  
 Percent Similarity: 94.84 Conservative: 1  
 Best Local Similarity: 94.78 Mismatches: 18  
 Query Match: 75.24 Indels: 12  
 Gaps: 1  
 DB: 8  
 US-10-773-302-2\_COPY\_1\_699 (1-699) x US-11-266-748A-319803 (1-1803)  
 QY 91 PheProThrGlnValIleProLeuLysThrLeuAsnAlaValAlaSerValProIleMet 110  
 DB 1722 TTCCGACACAGATCATCTCCATTAAAGCTGGAATGCAATGCTTCACTACCTAATG 1663  
 QY 111 TysSerTyrSerProLeuGlnGlnAsnPheMetValGluAspGluThrValLeuHisAsn 130  
 DB 1662 TATTCCTGTCTCCCTTACACAGAAATTTTATGTTGAAGATGAAACTGTTTACATNAC 1603  
 QY 131 IleProTyrMetGlyAspGlnValLeuAspGlnAspGlyTyrPheIleGluGlnLeuIle 150  
 DB 1602 ATTCCTTATATGAGAGATGAAATTTTATGATCAGATGATGACTTTCATTTGAAGAACTAATA 1543  
 QY 151 LysAsnTyrAspGlyLysValHisGlyAspArgGluCysGlyPheIleAsnAspGluIle 170  
 DB 1542 AAAATTTATGATGGGAAAGTACACCGGATGAGAAATGCGTTTATTAATGATGAATTT 1483  
 QY 171 PheValGluLeuValAsnAlaLeuGlyGlnTyrAsnAspAspAspAspAspAspGly 190  
 DB 1482 TTTTGAGAGTTGGGGAATGCCCTTGTCATATANNNNNNNNNNNNNNNNNNNNNNNGA 1423  
 QY 191 AspAspProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 210  
 DB 1422 GACGATCTGTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1363  
 QY 211 GluSerArgProProArgLysPheProSerAspLysIlePheGluAlaIleSerSerMet 230  
 DB 1362 GAAGCGCGCCCACTCGGAATTTCTTGTATTAATTTTGAAGCAATTTCTCTCAATG 1303  
 QY 231 PheProAspLysGlyThrAlaGluGlnLeuLysGlyLysTyrGlyGluLeuThrGluGln 250  
 DB 1302 TTTCCAGATTAAGGCAACGACGAAAGAACTAAAGAAATTAAGAACTCAACGAAACG 1243  
 QY 251 GlnLeuProGlyAlaLeuProProGluCysThrProAsnIleAspGlyProAsnAlaLys 270  
 DB 1242 CAGCTCCGAGCGGACATTCCTCGAATGTACCCCAACATAGATGACCAATGCTAAA 1183  
 QY 271 SerValGlnArgGluGlnSerLeuHisSerPheHisThrLeuPheCysArgArgCysPhe 290  
 DB 1182 TCTGTTCAGAGAGCAAGCTTACCTTCCTTCACTGCTTTCCTGTGAGCGAGTGT 1123  
 QY 291 LysTyrAspCysPheLeuHisPhePheHisAlaThrProAsnThrTyrLysArgLysAsn 310  
 DB 1122 AAATATGATGCTCTTCTTACATCTTTTCATGCAACCCCACTATTAAGCGAGAAAC 1063  
 QY 311 ThrGluThrAlaLeuAspAsnLysPProCysGlyProGlnCysTyrGlnHisLeuGluGly 330  
 DB 1062 ACAGAAACGCTCTAGCAACAACTTGTGACCAACAGTGTTCACAGCATTTTGAAGGA 1003  
 QY 331 AlaLysGluPheAlaAlaAlaLeuThrAlaGluArgIleLysThrProProLysArgPro 350

Db	1002	GCAAAGAGAGTTTGCTGCTGCTCTCAACCGGTGAGCGGATTAAGACCCCAACAAAGCTCCA	943
Qy	351	GIYGIYAARGARGAGIYARGLEUPROAANAENSERSEARGPROSETHPROTHRIE	370
Db	942	GGAGGCCCCAGAAAGAGGACCGGCTTCCCAATTAACATGACGAGCGCCAGACCCCAACAT	883
Qy	371	AAENVLLIENGUSERLYASPHTRHAPSEARAPARGIUALGILYTHGUTHNGIYGLY	390
Db	882	AATGTCGTGAATCAAAAGATACAGACATGATAGGAAGCAAGGACTGAACCGGGGGA	823
Qy	391	GIUAANAANAAPLYSGIUGIUGIUGIULYSLYASPGIUTHSERSESERSEGIUALA	410
Db	822	GAGAACATGATTAAGAGAGAGAGAGAGAAAGATGAACCTTCGAGCTCTCTGAAGCA	763
Qy	411	AANSEARARGYEGILNTHRPROLIELYSELTYPROASNIILEGIUPROPIGUAENVAL	430
Db	762	AATTCCTGGGTGTCAAACACCAATTAAGATGAAGCCAAATATATGAACTCTCTGAATGTG	703
Qy	431	GIUTPSPSEGIYALAGIULIASERMETPHEARGVALLEIILEGIYTHRYTYRABAPAN	450
Db	702	GAGTGAAGTGGTGCTGAACCTCTCAAGTTTAGAGTCTCATGTGGCACTTACTAGACAT	643
Qy	451	PHECYBALAILEALARGLEULIEGIYTHRYSTHRCYASRGILNVALTRYGIUNPEAR	470
Db	642	TTCTGTGCATTGCTGATAGTTAATTTGGAGCAAAACATGTATGACAGGTATGAGTTTGA	583
Qy	471	VALIYVGLUSERSETLIELALAPOLAPROLAIGIUNSPVALAPHTPROPOLARG	490
Db	582	GTCAAAGAGATCTAGCATCTAGCTCCAGCTCCCGCTGAGAGATGTGATATCTCTCCAAAG	523
Qy	491	LYSLVLYASXGLYSHISARGLEUTRPAALAHISCYASRGYSLILEGIN-LEULYSLY	510
Db	522	AAAAAGAGAGAAACACCGGTTGTGGCTGCACTGCAGAAAGATACAGCTGAAAAA	463
Qy	510	SAAPGLYSERSEASNIISVALTYRANTRYGINPROCYASAP-NISPROARGIAPROC	530
Db	462	GGACGGCTCTCTAACCATGTTTACAACATATCAACCTGTGATTCAATCCACGGCAAGCTT	403
Qy	530	YASAPSERSERCYAPROCYSVALILEALAGIUNAPHECYSGIULYSPHECYGILNYS	549
Db	402	GTCACAGTTCGTGGCCCTGTGTGATGACCAAAATTTTGGTAAAAAGTTTGTCAATGT	343
Qy	550	SERSERGIUCYBGLINAPNARG-PHAPROGLYCYASRCYSLYBALAGINCYASANTHL	569
Db	342	ACTTCAGAGTGCANAACCGCTTCCCGGATGCGGCTGCANAAGCAAGTGCACCA	283
Qy	569	YSGILNYSAPROCYSTYR-LEUALAVALARGIUCYASAPPROASPLEUCYSLEUTHRCYS	588
Db	282	AGCAGTGCCCGTGCTACCTCGGCTGTCCGAGAGTGTGACCTGACCTGTGTCTTACTTGT	223
Qy	588	GIYALIALAASPHIETTRAPSERLYASENVALISERCYELYASANCYASERTILEGINARG	608
Db	222	GGAGCGCGCTGACCATTTGGGACAGTAAAAAGTGTCTGCAGAAACATGCAATATTACCGG	163
Qy	609	GIYSERLYASHISLEULEULEUALAPROSEASPYAL-ALAGIYTRGILYILEPHEIL	628
Db	162	GGCTCCAAAAGCATTTATTTCTGGGCAACATCTGACGTGGCGAGGCTGGGGGATTTTAT	103
Qy	628	EIYASAPPROVALGINLYASANGIUNPHE--ILESERGIUNTRYCYSGIYULIIELE--	646
Db	102	CAAAAGTCTCTGTCAGAAAAATGAATTTTCATCTTCAGAAATACGTGGGAGATTAATTTT	43
Qy	647	----SERGINAPSGIUALAASPARGARGIYLYRVALTYR	658
Db	42	CTNCAAGGTGGAAGTGGACAGATGAGGGGANAAGTGTAT	3

RESULT 8  
US-11-266-748A-62177  
; Sequence 62177, Application US/11266748A  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Hartin, Paul  
; APPLICANT: Johnston, Patrick

```

APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcription Microarray Technology and
TITLE OF INVENTION: Methods of Using the Same
FILE REFERENCE: 5815-0102 (319189)
CURRENT APPLICATION NUMBER: US/11/266,748A
CURRENT FILING DATE: 2005-11-03
PRIOR APPLICATION NUMBER: EP 04105479.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105482.6
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105483.4
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105507.0
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105485.9
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105484.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: US 60/662,276
PRIOR FILING DATE: 2005-03-14
PRIOR APPLICATION NUMBER: US 60/700,293
PRIOR FILING DATE: 2005-07-18
NUMBER OF SEQ ID NOS: 483996
SOFTWARE: PatentIn version 3.3
SEQ ID NO 62177
LENGTH: 4606
TYPE: DNA
ORGANISM: Homo Sapiens
US-11-266-748A-62177

Alignment Scores:
Pred. No.:      1,366-191      Length:      4606
Score:          2386.50        Matches:     449
Percent Similarity: 75.6%       Conservative: 89
Best Local Similarity: 63.1%    Mismatches: 121
Query Match:    62.7%         Indels:     53
DB:             8              Gaps:       10

US-10-773-302-2 COPY_1_699 (1-699) x US-11-266-748A-62177 (1-4606)

QY      15 TTPARGTAAAGVVallySerglunTYrMetArGLueArgGlnleuluYaRPhreaArgc 34
Db      143 TGGAAAAGAAAAAGTAATAATCTGAATACATNGCGACTTCGCAACAATTAAACGGCTTGACGCA 202
QY      35 ALaasrgluVallySermetPheSerSerSerAanaTgGlnlyeIlleuGludrghTrglu 54
Db      203 AATATGGGTGCNAAGGCTTTGTATGTGGCNAATTTTGGCAAGAAGTTCMAAAGAACCCAG 262
QY      55 IlleuaanglngluTrrlyeglnlrgragtlleglnprovalHlerleuthrserVal 74
Db      263 ATCTTCATGAAGAAATGGAAGAAGCTTCGTGTCCAACCCTGTTCAGTCATTAAGACCTGTG 322
QY      75 SerSerLeuArgGlyThraArgGluCySerValThrSerAspLeu---AspPheProthr 93
Db      323 AGTGAACACCCCTTTCTGAAAAGGTACNATAGAGAGCATTTTCCGGGATTTGCAAGC 382
QY      94 GlnValIleProLeuylsrThreulanalaValAlaserValProIlemectryTSerTrp 113
Db      383 CAACATATGTTAATAGAGGTCACTGAACAACAGTTGCAATGGTTCCTCCATCATGTATTCCTGG 442
QY      114 SerProLeuGlnGlnAsnPhmetValGluuaRpgluThrValleuHlsanleProtyr 133
Db      443 TCCTCTCTCCCAACAACCTTATATGGTAGAAGATGAGACGGTTTTGTGCATATATCTCCCTAC 502
QY      134 MetGlyaaRpgluValleuanRpglnAspGlyThrPheIleGlnGluLeuileuysaenTy 153
Db      503 ATGGAGATGAAGTGAAGAAACAAGATGACATTTTATTGAGAGCTGATCATTAAGTAT 562
QY      154 AspglyLyVaIHleGlyAspArgGlu-----CysGlyPheIleAnaRpglu 169
Db      563 GATGGGAAGATCAATGCTGTAAGAAGAGATGATCCCTGAGATCCGTTCTGATTAAGTATGCT 622
QY      170 IlepheValGluLeuValaenlaAleuGlyGlnTyfAsnAspAspAsp----- 185

```



[illegible]

Db	1635	CTGAAAGAAAGTAATCTTCCACACAGTGTACAACTACCAACCCCTGGACCAACCAAC	168
Qy	528	GlnProCysAbpSerSerCysProCysVal11lealaginPheCysGlnuPheCys	547
Db	1685	CGCCCTGTGTACACACCTGCCCTCCGTGATCTGACTCAGAAATTTCTGTGAAATGTTCTGC	1744
Qy	548	GlnCysSerSerGluCysGlnInaPhePheProGluCysArgCysLysAlaGlnCysAsn	567
Db	1745	CAGTGCAACCCAGACTGTGAGAAATCGTTTCCCTGAGCTGTGCTGTAAAGACCAAGTGCAT	1804
Qy	568	ThirysGlnCysProCysTyTrLeuAlaValArgGluCysAbpProAbpLeuCysLeuThr	587
Db	1805	ACCAAGCAATCTCTTGTCTATCTGGCAGTGGAGAAATGTGACCTGACCTGTGTCTAC	1864
Qy	588	CysGlyAlaAlaAAPHI:STRPASPserLysaenValSerCysLysaenCysSerTleGln	607
Db	1865	TGTGGGGGCTCAGAGCACTGGGACTGCAGAGGTGGTTCCTGTAAAACTGCAGCATCCAG	1924
Qy	608	ArgGlySerLysLysHleuLleuLleuLleuAlaProSerAspValAlaGlyTrpGlyTlePhe	627
Db	1925	CGTGACTTAAAGAAAGCACCTGTCTGTGCCCCCTGTATGTGGCCGAGATGGGACCTTC	1984
Qy	628	IleLysAbpProValGlnLysAsnGluPheIleSerGluTyCysGlyGlnIleIleSer	647
Db	1985	ATTAAGAGAGTCTGTGCAGAAAGACGAATTCATTTCTGAATCTGTGGAGCTCATCTCT	2044
Qy	648	GlnAbpGlnAlaAEPATGATgGlyLysValTyraAspLysTyTrMetCysSerPheLeuPhe	667
Db	2045	CAGATGAGGCTGATGCACCGCGAAAGGTCTATGACAAATCTCATGTCCAGCTCTCTCTTC	2104
Qy	668	AsnLeuAsnAsnAspPheValValAspAlaThrArgLysGlyAsnLysLleArgPheAla	687
Db	2105	AACTCAATATATGATTTTGTATGTGATGCTACTGGAAAGGAAACAAATTCGATTTGCA	2164
Qy	688	AsnHisSerValAsnProAsnCysTyTrAlaLysVal	699
Db	2165	AATCATTCAGTAATCCCACTGTTATGCCAAAGTG	2200
RESULT 9			
US-11-266-748A-26094			
Sequence 26094, Application US/11266748A			
Publication NO. US20060134663A1			
GENERAL INFORMATION:			
APPLICANT: Hartin, Paul			
APPLICANT: Johnston, Patrick			
APPLICANT: Mulligan, Karl			
TITLE OF INVENTION: Transcriptome Microarray Technology and			
TITLE OF INVENTION: Methods of Using the Same			
FILE REFERENCE: 55815-0102 (319189)			
CURRENT APPLICATION NUMBER: US/11/266, 748A			
CURRENT FILING DATE: 2005-11-03			
PRIOR APPLICATION NUMBER: EP 04105479.2			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: EP 04105482.6			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: EP 04105483.4			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: EP 04105507.0			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: EP 04105485.9			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: EP 04105484.2			
PRIOR FILING DATE: 2004-11-03			
PRIOR APPLICATION NUMBER: US 60/662, 276			
PRIOR FILING DATE: 2005-03-14			
PRIOR APPLICATION NUMBER: US 60/700, 293			
PRIOR FILING DATE: 2005-07-18			
NUMBER OF SEQ ID NOS: 48396			
SOFTWARE: PatentIn version 3.3			
SEQ ID NO 26094			
LENGTH: 4640			
TYPE: DNA			
ORGANISM: Homo Sapiens			





```

Db 749 TAATTTCTTGTGTGTTTTTGACAGGTTGTGGGCTGCACACTGCAGAAAGTACAGCTGAA 690
Qy 510 Lys ----- 510
Db 689 AAGGGTAGCATCTTCCATTCCTCTCATTAATTAGCTTAACAATATCTGCTTGTTCGA 630
Qy 510 ----- 510
Db 629 TTTGTTTGAATAGTAAGTGAATAGAGCTGGGTTAATTCTAATTAATGCTTCTGT 570
Qy 510 ----- 510
Db 569 ATGTTAAACCAATTGGAATGAGAGAGCTCTACTGATGAATCCCAAGTCAATTCTTCG 510
Qy 510 ----- 510
Db 509 TAAAGTGAGGCTGACTGGAAGGTGTGATCCGAGTCAATTGAAGGAAGAGTCTTGGG 450
Qy 510 ----- 510
Db 449 ATGGAGTGAAGTGTAGCCGCTCTTATAGTAATCATTAAGCACTTGGGGGTGGAGAG 390
Qy 510 ----- 510
Db 389 TAATTTTGCATCGTTGTGGAACCGGAATATGTGTAAATGTATGCATTCAGTAAGGC 330
Qy 510 ----- 510
Db 329 AGGCAGAGTACACTCCACAGGTAGTAGGAAGAATGACTCGTAAATTGTATGATC 270
Qy 511 ----- 514
Db 269 GTTTCATCTCCCTGATTCATATGCGCTCAGATGTTACTTACTCAGACGGCTCTC 210
Qy 514 TASHHtstValTYTAgmTYrgLnProCYsAaBpHtspRoAgrLnProCYsAaBpserCy 534
Db 209 TAACCATGTTTAACTATCAACCCGTATCATCCACGGACGCTTGTGACGTTGCGT 150
Qy 534 sProCYsValIIEaLagLnAaBpHeCYsGluYsPheCYsGlnCYsSerSerGluCYsG 554
Db 149 CCCTGTGTGATGACCA-ATTTTGTGTGAAAAGTTTGTCAATGTAGTTCAAGAGTGA 91
Qy 554 nAsnArGpHeProGlyCYsArGcYs 562
Db 90 GTATTGTGCTTGTGATGCATTTGC 66

RESULT 11
US-11-266-748A-241375
; Sequence 241375, Application US/11266748A
; Publication No. US2006013463A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276

```

[illegible]

QY 446 ThTyrTrpAspPheCysAlaIleAlaArgLeuIleGlyThrLysThrCysArgGln 465  
 Db 786 ACTTACTATGACAAATTTCTGTGCTTCTGTAGTTAATTTGGACCAAAACATTTAGACAG 845  
 QY 466 ValTyrGluPheArgValIleGluSerSerIleIleAlaProAlaProAlaGluAspVal 485  
 Db 846 GTGATGATGTTTAAAGTCAAGAAATCTAGCATCATAGCTCCAGCTCCGCTGAGATGTG 905  
 QY 486 AspThrProProAlaGlyLysLysLysArgLysHis 496  
 Db 906 GATACTCTCCACAGAAAAAGAGAAACACCGCTTTAGAGATCTTCTGAATTT 965  
 QY 496 ----- 496  
 Db 966 TGTGTAATTAATTTCTGACCAAGTACTGCTGTGTTAGATGCTGAAATTAAGA 1025  
 QY 496 ----- 496  
 Db 1026 GGAAGTATCATGTGTCATCATGAGAAATTAATCAAGGACAGACACCTGTCTTT 1085  
 QY 496 ----- 496  
 Db 1086 GTGAGCATTTGGAGCTCAGAGTTCTGACAGAGATGATGATGAAACAGAGAAATGAC 1145  
 QY 496 ----- 496  
 Db 1146 GTCTTGCCACACTGGGGAGCTGAGCCATCAAGCTGTGTTGAGTTGCTTCAACTGATG 1205  
 QY 496 ----- 496  
 Db 1206 GCAAAATAAATCTGGGAGGAGCATGACTTTGGCTTTAAACCATTTCTGTATTCATGCA 1265  
 QY 496 ----- 496  
 Db 1266 CAGATAGATCCAGTTAAGCAGCTCTCTGTGGAATTTGAGCTTCCCGAGAAATTTGTT 1325  
 QY 497 ----- 497  
 Db 497 ----- ArgLeuThrAlaIleAlaHisCysArgLysIleGluLeuLys 509  
 Db 1326 TAAATTTCTTTGTTTGGAGTTGGCTGACACTGAGAAATATCAGCTGAAA 1385  
 QY 510 ----- 510  
 Db 1386 AAGGTTAGCATCTTTCATCTCTCATTTATTAAGCTTAACAAATCTGTTTGTTC 1445  
 QY 510 ----- 510  
 Db 1446 TTTGTTTAAAGTAACTGATGATGATGAGCTTGGGTTATTTAAATTAATTTGCCTTCTGT 1505  
 QY 510 ----- 510  
 Db 1506 ATGTTAAACCAATTTGATGAGAGAGAGCTCTATGATGAAATCCAGTCATTTCTTCTGT 1565  
 QY 510 ----- 510  
 Db 1566 TAAAGTGAGGCTGACTGGAAGGTGTGAGTCCAGTCACTTGAAGAGAAATGCTTTGGG 1625  
 QY 510 ----- 510  
 Db 1626 ATGGGACTGAGATGAGCCGTGCTTTAGATGAACTATAAGCACTTGGGGGTGGAGAG 1685  
 QY 510 ----- 510  
 Db 1686 TATTTTTCATGCTGTGTGTGAAAACCGGAATATGTATATGTTATGCAATTCATTAAGGC 1745  
 QY 510 ----- 510  
 Db 1746 AGGCCAGCTACCTCCACAGGTAATGAGAAAGATTAATCTGTGTTAATTTGTATATATC 1805  
 QY 511 ----- AspGlySerSe 514  
 Db 1806 GTTTCATCTCCCTGATTCATGATGAGCTGATGATGATTAATCTTAAGTCAAGCGCTCTC 1865  
 QY 514 rAnHisValTyrLeuTyrGlnProCysAspHisProAlaGlnProCysAspSerSerCys 534

Db 1866 TAAACATTTAGACATATGACCTGTGATCATCCAGGAGCTTGTGACAGTTGCTG 1925  
 QY 534 sProCysValIleAlaGlnAspPheCysGluLysPheCysGlnCysSerSerGluCysGln 554  
 Db 1926 CCTTGTGTGTATGACACA-AATTTTGTGAAGAGTTTGTCAAGTATGTTGACAGTGTAA 1984  
 QY 554 nAsnArgPheProGlyCysArgCys 562  
 Db 1985 GTATTTGTTGCTTTGATGCAATTTGC 2009  
 RESULT 12  
 US-11-266-748A-98588/C  
 ; Sequence 98588, Application US/11266748A  
 ; Publication No. US2006013463A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harlin, Paul  
 ; APPLICANT: Johnston, Patrick  
 ; APPLICANT: Mulligan, Karl  
 ; TITLE OF INVENTION: Transcription Microarray Technology and  
 ; FILE REFERENCE: 55815-0102 (319189)  
 ; CURRENT APPLICATION NUMBER: US/11/266, 748A  
 ; CURRENT FILING DATE: 2005-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105479.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105482.6  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105483.4  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105507.0  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105485.9  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: EP 04105484.2  
 ; PRIOR FILING DATE: 2004-11-03  
 ; PRIOR APPLICATION NUMBER: US 60/662,276  
 ; PRIOR FILING DATE: 2005-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/700,293  
 ; PRIOR FILING DATE: 2005-07-18  
 ; NUMBER OF SEQ ID NOS: 483996  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 98588  
 ; LENGTH: 1140  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (86)..(86)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (169)..(169)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (367)..(367)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (468)..(468)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; US-11-266-748A-98588  
 Alignment Scores:  
 Pred. No.: 1.55e-141 Length: 1140  
 Score: 1785.00 Matches: 350  
 Percent Similarity: 92.5% Conservative: 6  
 Best Local Similarity: 90.9% Mismatches: 21

Query Match: 46.9% Indels: 10  
DB: 8 Gaps: 0  
US-10-773-302-2\_copy\_1\_699 (1-699) x US-11-266-748A-98588 (1-1140)  
Qy 274 ArgGluGlnSerLeuHisSerPheHisThrLeuPheCysArgGlyCysPheLeuTyrAsp 293  
Db 1138 AGAAGGCAAGAGCTTACCTCTTTCATACGCTTTCTGTAGGCAATGTTTAAATATGAC 1079  
Qy 294 CysPheLeuHisProPheHisAlaThrProAsnThrTyrLeuAspGlyAsnThrGluThr 313  
Db 1078 TGCTTCTTCACTCTTTTCATGCAACCCCAACCTTTATAGCCGAGAAACACAGAAACA 1019  
Qy 314 AlaLeuAspAsnLeuProCysGlyProGlnCysTyrGlnHisLeuGluValAlaGlu 333  
Db 1018 GCTCTAGACCAAACTTGTGACACACAGTGTATCCGACATTTGAGAGGAGCAAGAG 959  
Qy 334 PheAlaAlaAlaLeuThrAlaGluArgGlyLeuThrProProLeuAspProGlyValArg 353  
Db 958 TTTCTGTCTCTCAACCGCTGACGGGATTAAGACCCCAACCAACCTCCAGAGGCGC 899  
Qy 354 ArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu 373  
Db 898 AGAAGAGGACGGCTTCCCAATACATGACAGCCCAACCCCAACCTTATATGTCTG 839  
Qy 374 GluSerLeuAspThrAspSerAspArgGluAlaGlyThrGluThrGlyValAsnAsn 393  
Db 838 GAATCAAAAGATACAGACAGATGAGGACAGGACCTGAACCGGGGGAGAGAACAT 779  
Qy 394 AspArgGluGluGluGluValLeuAspGluThrSerSerSerSerGluAlaAsnSerArg 413  
Db 778 GATAAAG 719  
Qy 414 CysGlnThrProIleLeuMetLeuProAsnIleGluProProGluAsnValGluTyrSer 433  
Db 718 TGTCAACACCAATTAAGATAGAGCAATTTGAACTCTTGAGATGTGAGTGTGAGT 659  
Qy 434 GluValAlaGluAlaSerMetPheArgValLeuIleGlyTyrTyrAspAsnPheCysAla 453  
Db 658 GGTGCTGAGGCTCAATGTTTAAAGTCTCAATTTGACCTTCAATGACATTTCTGTGTC 599  
Qy 454 IleAlaAlaGluLeuIleGlyThrLeuThrCysArgGlnValTyrGluThrPheArgValGlu 473  
Db 598 ATTGTAGATTAAATTGGACCAAAACATGTAGACAGGTATGAGTTTAAAGTCAAAATA 539  
Qy 474 SerSerIleIleAlaProAlaProAlaGluAspValAspThrProProArgLeuLeuLeu 493  
Db 538 TCTAGCATATAGCTCCAGCTCCGCTGAGATGTGATCTCTCCCAAGAAAGAAAGAG 479  
Qy 494 ArgLeuHis-ArgLeuThrAlaAlaHisCysArgGlyValLeuLeuLeuAspGlySe 513  
Db 478 AGAAGACACGNGTGTGGGTGACACTGACAGAAAGATACGCTGAGAAAGACGGCTC 419  
Qy 513 rSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnPro-CysAsp-Ser 532  
Db 418 CTCTAACCATGTTTAACTATCAACCTGTGATCATCAAGGACCTTGTATGACNAGT 359  
Qy 533 SerCysPro-CysValIleAlaGlnAsnPheCys--GluLeuPheCysGlnCysSerSer 551  
Db 358 TGGAGCCCTTGTGTGATGACCAAAATTTTGTGTAAGATTTTGTGCATATGATTTCA 299  
Qy 552 GluCysGlnAsnArgPheProGlyCysArgCysLeuAlaGlnCysAsnThrLeuGlnCys 571  
Db 298 GAGTGTCAAAACCGCTTTCCGGGATGCGGCTGCAAAAGACAGTGCACACCAAGACAGTC 239  
Qy 572 ProCysTyrLeuAlaValAlaArgGluCysAspProAlaPleuCysLeuThrCysGlyAlaAla 591  
Db 238 CGGTCTACTGCTGGCTGTCCGAGAGTGTGACCTGTGACTCTACTGTTGGAGCGCT 179  
Qy 592 AspHisTyrAspSerLeuAsnValSerCysLeuAsnCysSerIleGlnArgGlySerLeu 611  
Db 178 GACCATGTGACATGATTAATGTGTCTGCAAGAACTGACATTTCAAGCGGGCTCCAA- 120

Qy 612 LysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIle-PheIleLysAsp 631  
Db 119 AACCATCTATGTCTGACCATCTGACTGGAGAGCTTGGGGATTTTATTCAGAGTCC 60  
Qy 631 oValGlnLysAsnGluPheIleSerGluTyrCysGlyGluIleIle-SerGlnAspGlu 650  
Db 59 TGGGCAAGAAATGAGATT-CTCTCAAAATCTGGGGAGAAATTTTCTCCAGATGAA 3  
RESULT 13  
US-11-266-748A-151399  
Sequence 151399, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Hartkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Malligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 48396  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 151399  
LENGTH: 1140  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (673)..(673)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (774)..(774)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (779)..(779)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (972)..(972)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1055)..(1055)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-151399  
Alignment Scores:  
Pred. No.: 1.55e-141 Length: 1140  
Score: 1785.00 Matches: 350  
Percent Similarity: 92.5% Conservative: 6  
Best Local Similarity: 90.9% Mismatches: 21  
Query Match: 46.9% Indels: 10  
DB: 8 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-11-266-748A-151399 (1-1140)

QY 274 ArgGluGlnSerLeuHisSerPheHisThrLeuPheCysArgArgCysPheLeuYTrpAp 293  
Db 3 AAGAGGCAAGGCTTACACTCTTTCATAGCGCTTTCCTGAGGAGGATTTTAAATATGAC 62

QY 294 CysPheLeuHisProPheHisAlaThrProAsnThrTyrLeuArgLeuHisThrGluThr 313  
Db 63 TCGCTTCTTCACTCTTTCATAGCAACCCCACTTATAGCGAGAGAGACAGAAACA 122

QY 314 AlaLeuAspAsnLeuAspProCysGlyProGlnCysTyrGlnHisLeuGluGluValAlaYsglu 333  
Db 123 GCTTATAGACAAACACTTGTGGACCAAGTGTATACACATTTGGAGGAGAGCAAGAG 182

QY 334 PheAlaAlaLeuThrAlaGluArgLeuThrProProLeuArgProGlyGlyValArg 353  
Db 183 TTGTGCTGCTCTCAACCGCTGAGCGGATTAAGACCCCAACAAAGTCCAGAGGCCCC 242

QY 354 ArgArgGlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeu 373  
Db 243 AAGAGAGGACGGCTTCCCAATTACAGTACAGGCCCAAGCCCACTTATATGTCTG 302

QY 374 GluSerLeuAspThrAspSerAspArgGluAlaGlyThrGluThrGlyGlyGluAsn 393  
Db 303 GAATCAAGGATACAGACAGTATGGAGAGCGGACAGAAACGGGGGAGAGACAT 362

QY 394 AspLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 413  
Db 363 GATTAAG 422

QY 414 CysGlnThrProIleLeuSerLeuProAsnIleGluProProGluAsnValGluTrpSer 433  
Db 423 TCTCAACACCAATTAAAGATGAGACCAATATTGAACCTCTGAGATGTGGAGT 482

QY 434 GlyAlaGluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAla 453  
Db 483 GGTGCTGAAGCCTCAATGTTTGAAGTCCATTTGGACATTAATAGACATTTCTGTGC 542

QY 454 IleAlaGluLeuIleGlyThrTyrThrCysArgGlnValTyrGluPheArgValYsglu 473  
Db 543 ATTGCTAGGTTAATTGGGACCAAAACATGTACAGGTGTATGAGTGTGACGTCAAGAA 602

QY 474 SerSerIleIleAlaProAlaProAlaGluAspValAspThrProProArgLeuYsglu 493  
Db 603 TCTACATCATATGCTCCAGCTCCCGTAGAGATGTGATCTCCCAAGGAGAAAGAG 662

QY 494 ArgLeuHis-ArgLeuTrpAlaAlaHisCysArgGlyIleGlnLeuYsglyAspGlySe 513  
Db 663 AGGAACAACCGGTTGTGGGCTGCACATCGCAGAAAGATACAGCTGAGAAAGAGCGCTC 722

QY 513 rSerAsnHisValTyrAsnTyrGlnProCysAspHisProArgGlnPro-CysAsp-Ser 532  
Db 723 CTCTAACCATGTTTCAACTATCAACCCCTGTATCATCAACCGCAGCTGTGTCAGTCA 782

QY 533 SerCysPro-CysValIleAlaGlnAsnPheCys--GluYsgPheCysGlnCysSerSer 551  
Db 783 TGTGTCCTTGTGTGATAGCAAAATTTTGTGTAAATTTTGTCAATGTATGTTCA 842

QY 552 GluCysGlnAsnArgPheProGlyCysArgCysLeuAlaGlnCysAsnThrLeuYsglinCys 571  
Db 843 GAGTGTCAAAACCGCTTTCGGGATCGCGCTGCAGAGCAAGTGCAGACCAAGCAGTGC 902

QY 572 ProCysTyrLeuAlaValArgGluCysAspProAspLeuCysLeuThrCysGlyAlaAla 591  
Db 903 CCGTGTCTACTGCTGCTCGAGAGTGTGACCTGTGACCTCTGTCACTGTGTGAGCGCT 962

QY 592 AspHisTrpAspSerLeuAsnValSerCysLeuAsnCysSerIleGlnArgGlySerLeu 611  
Db 963 GACCATTTGAGACAGTAAATATGTCTGCAAGATGTGAGTATTCAGGGGGGCTCCAA- 1021

QY 612 LysHisLeuLeuLeuAlaProSerAspValAlaGlyTyrGlyIle-PheIleLeuAspPr 631  
Db 1022 AAGCATCTATTTGCTGACATCTGACTGGAGAGGCTTGGGGGATTTTATTCAGAGATCC 1081

QY 631 oValGlnYsgAsnGluPheIleSerGluTyrCysGlyGluIleIle-SerGlnAspGlu 650  
Db 1082 TGGGAGAGAAATGATGAT--CTCTCAAAATACGTGGGGGAGATATTCTCCAGATGAA 1138

RESULT 14  
US-11-266-748A-175796  
Sequence 175796, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 175796  
LENGTH: 1011  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (18)..(18)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (64)..(87)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (141)..(141)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (151)..(151)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (212)..(212)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (222)..(222)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (238)..(238)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (251)..(251)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-266-748A-175796



## Alignment Scores:

Pred. No.: 1.55e-120 Length: 1011  
Score: 1536.00 Matches: 315  
Percent Similarity: 92.7% Conservative: 3  
Best Local Similarity: 91.8% Mismatches: 14  
Query Match: 40.3% Indels: 11  
DB: 8 Gaps: 0

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-11-266-748A-175796 (1-1011)

```
QY 162 GUCYGGIYPhelIeaaspGluIlePheValGluLeuValAspAlaLeuGlyGln-Ty 181
DB 2 GAATGTGGGTATTAATATGATGATATTTTGGGTGGTGAATGCGCTTGGCAATA 61
QY 181 rAspAspAspAspAspAspAspAspGlyAspAspProGluGluArgGluGluValGly 201
DB 62 TANNNNNNNNNNNNNNNNNNNNNNNGAGACATCTGAGAGAAAGAAAGAAAGCAAGA 121
QY 201 sAspLeuGluAspHisArg-AspAspLys-GluSerArgProPro-ArgLysPhePro-S 220
DB 122 AGATTGTGAGGATCCCGCAGATGATTAANAGAAAGCCGCCACCTCGGAATTTCTCT 181
QY 220 eraAspLysIlePhe-GluAlaIleSerSer-MetPhePro-AspLysGlyThrAla-Glu 238
DB 182 CTGATAAATTTTGGAGCCATTTCTCANATGTTTCCANGATTAAGGCGCACAGCANGAA 241
QY 239 GluLeuLys-GluLysTyrlYsGluLeuThrGluGluGlnLeuProGluValAlaLeuPro 258
DB 242 GAACATAAGNGAAATATAAGAACTCACCGAACGAGCTCCAGGGGCACTTCTCTCC 301
QY 258 OGluCyThrProAsnIleAspGlyProAsnAlaLysSerValGlnArgGluGlnSerLe 278
DB 302 TGAATGTAACCCCAACATATGATGAGCAAACTCTAACTGTTTGAGAGAGCAAAAGCTT 361
QY 278 uHisSerPheHisThrLeuPheCyArgArgCyPheLysTyrlYsArgCyPheLeuHisPr 298
DB 362 ACACCTCTTTCATACGCTTTTCTGTAGGCGATGTTTAAATATGACTGCTCTTCAATCC 421
QY 298 oPheHisAlaThrProAsnThrTyrlYsArgLysAsnThrGluThrAlaLeuAspAspLys 318
DB 422 TTTTATGCAACCCCAACATATGATGAGGAGAAACACAGAAACAGCTCTTGAACAA 481
QY 318 sProCyGlyProGlnCyTyrlYsGlnHisLeuGluValAlaLysGluPheAlaAlaAla 338
DB 482 ACCTGTGGAACCAAGTGTACACGATTTGGAGGAGAAAGAGTGTGCTGCTGCTCT 541
QY 338 uThrAlaGluArgIleLysThrProProLysArgProGluGlyArgArgGlyValArgLe 358
DB 542 CACCGCTGAGCGGATTAAGACCCCAACAAACGTCACAGAGGCGCCACAGAAAGAGCGCT 601
QY 358 uProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSerLysAspTh 378
DB 602 TCCCAATTAACATAGAGGCGCCAGCACCCCACTTAATGTGCTGGAATTCAAAGAGATC 661
QY 378 rAspSerAspArgGluAlaGlyThrGluThrGlyGluAsnAsnAspLysGluGluGlu 398
DB 662 AGACAGTGTATAGGAAGCAAGGAGCTGAAACGGGGGAGAGAAACAATATTAAGAGAGA 721
QY 398 uGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCyGlnThrProI 418
DB 722 AGAGAAAGCAAGATTAACCTTCAGAGCTCTCTGAGCAAAATTCGCTGCAACACCAAT 781
QY 418 eLysMetLysProAsnIleGluProProGluAsnValGluTrpSerGlyAlaGluAlaSe 438
DB 782 AAAAGATGAAGCCAAATATTAACCTCTAGAGATGTGAGTGAAGTGGTCTGAAGCTTC 841
QY 438 rMetPheArgValLeuIleGlyThrTyrlYsAspAsnPheCyAlaAlaLeuArgLeuI 458
DB 842 AATGTTTAGAGTCCCTCATTTGGCACTTAATGACATATTTCTGTGCAATTTGCTAGGTAT 901
QY 458 eGlyThrLysThrCyArgGlnValTyrlGlu-PheArgValLysGluSerSerIleIleA 478
```

DB 902 CGGACCAAAACATGTAGACAGGTGTATGATTAAGATCAAGAAATCTAGCATCAG 961

QY 478 lAProAlaProAlaGluAspValAspThrProProArgLysLysLys 493

DB 962 CTCACGCTCCCGCTGAGAGATGTGATACTCTCCAAAGAAAGAAAGA 1008

## RESULT 15

US-11-266-748A-361225/c

Sequence 361225, Application US/11266748A

Publication No. US2006013463A1

GENERAL INFORMATION:

APPLICANT: Hartin, Paul

APPLICANT: Johnston, Patrick

TITLE OF INVENTION: Transcription Microarray Technology and

FILE REFERENCE: 55815-0102 (319189)

CURRENT APPLICATION NUMBER: US/11/266, 748A

CURRENT FILING DATE: 2005-11-03

PRIOR APPLICATION NUMBER: EP 04105479.2

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: EP 04105482.6

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: EP 04105483.4

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: EP 04105507.0

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: EP 04105485.9

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: EP 04105484.2

PRIOR FILING DATE: 2004-11-03

PRIOR APPLICATION NUMBER: US 60/662, 276

PRIOR FILING DATE: 2005-03-14

PRIOR APPLICATION NUMBER: US 60/700, 293

PRIOR FILING DATE: 2005-07-18

NUMBER OF SEQ ID NOS: 483996

SOFTWARE: PatentIn version 3.3

SEQ ID NO 361225

LENGTH: 904

TYPE: DNA

ORGANISM: Homo Sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (30)..(30)

OTHER INFORMATION: n is a, c, g, or t

FEATURE:

NAME/KEY: misc feature

LOCATION: (134)..(134)

OTHER INFORMATION: n is a, c, g, or t

FEATURE:

NAME/KEY: misc feature

LOCATION: (139)..(139)

OTHER INFORMATION: n is a, c, g, or t

FEATURE:

NAME/KEY: misc feature

LOCATION: (205)..(205)

OTHER INFORMATION: n is a, c, g, or t

FEATURE:

NAME/KEY: misc feature

LOCATION: (240)..(240)

OTHER INFORMATION: n is a, c, g, or t

US-11-266-748A-361225

## Alignment Scores:

Pred. No.: 1.71e-110 Length: 904  
Score: 1416.50 Matches: 280  
Percent Similarity: 93.1% Conservative: 2  
Best Local Similarity: 92.4% Mismatches: 11  
Query Match: 37.2% Indels: 10  
DB: 8 Gaps: 3

US-10-773-302-2\_COPY\_1\_699 (1-699) x US-11-266-748A-361225 (1-904)

```
OY 276 GlnSerLeuHisSerPheHisThrLeuPheCysArgArgCysPheLysThrAspCysPhe 295
DB 904 CAAAGCTTACACCTCTTTCATACGCTTTCTGAGCGAGATGTTTAAATATGACTCTTC 845
OY 296 LeuHisProPheHisAlaThrProAsnThrTyrLysArgLysAsnThrGluThrAlaLeu 315
DB 844 CTACATCCTTTTCATGCAACACCAACACTTATAGCGAAGAAACAGAAACAGCTCTA 785
OY 316 AspAsnLysProCysGlyProGlnCysTyrGlnHisLeuGluGlyAlaLysGluPheAla 335
DB 784 GACCAAAACCTTGAGACCAAGTGTACACAGCATTTGAGGAGCAAAAGAGTTTGCT 725
OY 336 AlaAlaLeuThrAlaGluArgGlyLeuSerProProLysArgProGlyArgArgArg 355
DB 724 GCTGCTCTACCCCTGAGCGGATTAAGACCCCAACCAACGTCACAGAGCGCGAGAAAG 665
OY 356 GlyArgLeuProAsnAsnSerSerArgProSerThrProThrIleAsnValLeuGluSer 375
DB 664 GGAAGGCTTCCCAATACAGTAGCAGGCCAGCACCCCAACATTAATGCTGGAATCA 605
OY 376 LysAspThrAspSerAspArgGlyAlaGlyThrGluThrGlyGlyGluAsnAsnLys 395
DB 604 AAGGATACAGACAGTATAGGAGGAGCAAGCACTGAAACGGGGGAGAGAAACAATGATAA 545
OY 396 GluGluGluGluLysLysAspGluThrSerSerSerSerGluAlaAsnSerArgCysGln 415
DB 544 GAAAGAAAGAAAGAAAGATGAACCTTGAGCTCTTGAAACAAATTCGGGTGCA 485
OY 416 ThrProIleLysMetLysProAsnIleGluProProGluAsnValGluTyrSerGlyAla 435
DB 484 AACACCAATAAAGATGAAGCAAAATATTGAACCTCGAAGATGTGAGTGGAGTGTGCT 425
OY 436 GluAlaSerMetPheArgValLeuIleGlyThrTyrTyrAspAsnPheCysAlaIleAla 455
DB 424 GAAGCTCATATGTTAGATGCTCATTTGGCAGCTTACTATGACAAATTCGTGCAATGCT 365
OY 456 ArgLeuIleGlyThrLysThrCysArgGlnValTyrGluPheArgValLysGluSerSer 475
DB 364 AGCTTAAATGGAGCAAAACATGTAGACAGGTGTATGAGTTTGAAGTCAAGAAATCTTAC 305
OY 476 IleIleAlaProAlaProAlaGluAspValAspThrProProArgLysLysArgLys 495
DB 304 ATCATAGCTCCAGCTCCGCTGAGATGTGATACTCTCCAAGGAAAGAAAGAGAGAA 245
OY 496 His-ArgLeuThrAlaAlaHisCysArgGlyValIleGlnLeuLysLysAspGlySerSerAs 515
DB 244 CACCGAGTTGTGGCTGCACACTGCAGAAAGATACAGCTGAAAGAGACGGCTCCTCTAA 185
OY 515 HisValTyrAsnTyrGlnProCysAspHisProArgGlnPro-CysAsp-SerSerCys 534
DB 184 CCAATGTTACAACTACAAACCTGTGATATCCACGCGACCTTGATGACNAGATTCGTGC 125
OY 535 Pro-CysValIleAlaGlnAsnPheCys--GluLysPheCysGlnCys---SerSerGlu 552
DB 124 CTTTGTGTAGATAGCAAAATTTTGTGTAAAGTTTGTCAATGATGATTTTCAGAG 65
OY 553 CysGln-AsnArg---PheProGlyCysArgCys---LysAlaGlnCysAsnThr 568
DB 64 TGTCAAAACCGCTTTTCCCGGAGTGCCTGTCNCAAAAGCACAGTGCAAAACA 10
```

Search completed: August 13, 2006, 16:58:50  
Job time : 323.344 secs

THIS PAGE LEFT BLANK